THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Determinants of Radiographic Severity in Ankylosing Spondylitis

Steven Truong

BSc, MBBS

A thesis submitted for the degree of Master of Philosophy at

The University of Queensland in 2015

The University of Queensland Diamantina Institute

ABSTRACT

Ankylosing Spondylitis (AS) is a highly heritable condition characterised by excessive bone formation and stiffness of the spine and pelvis causing functional impairment. Functional impairment is strongly correlated with radiographic signs of bone formation. Clinicians can accurately measure bone formation but have a limited ability to predict or prevent it.

A small number of non-genetic factors have been identified that explain approximately one third of the variation in radiographic severity. No genetic predictors of radiographic severity have consistently been identified.

This work investigates the contribution of non-genetic and genetic factors to radiographic severity. Knowledge of causal genetic variants could be used to predict prognosis or to design therapeutics. Spinal radiographs were scored using the modified Stoke Ankylosing Spondylitis Severity Score (or a modified version of this score), mostly by a single calibrated reader.

Observed modified Stoke Ankylosing Spondylitis Severity Scores were compared between males and females. Cross-sectional modelling was performed on 1392 subjects by comparing previously recorded clinical covariates and demographics with the latest radiographic score for each patient, using a negative binomial model. Residuals from this model were considered to be the phenotype corrected for known covariates. A quantitative Genome Wide Association Study (GWAS) was performed on 1253 cases, comparing the corrected phenotype with millions of genetic markers per subject. Genome-wide Complex Trait Analysis (GCTA) was used to estimate the contribution of common variants with small effect sizes to the phenotype. An allelic risk score approach was used to examine the hypothesis that known bone mineral density, height, AS susceptibility and smoking dependence related loci also affect AS radiographic severity.

Known contributors to radiographic severity - gender, smoking and symptom duration were found to explain approximately one-third of phenotypic variation. GCTA estimated that common variants contribute at least 33% of phenotypic variation. No signal was identified to link known contributors to AS susceptibility, including the MHC region, with radiographic severity. Known AS susceptibility, BMD, height and smoking dependence loci were not associated with severity upon allelic risk score analysis. GWAS suggested that loci involved in bone metabolism and innate inflammation may influence radiographic severity. Some genes in the same pathways as AS susceptibility loci, including the TNF pathway, had a suggestive association with the phenotype. Additionally suggestive associations between loci related to bone formation and immune-microbiota interactions in the gut were present, which, if confirmed, may provide a mechanism for observed patterns of microbiota in human AS cases.

This study did not definitively associate any specific loci with the phenotype, probably due to its low power and limitations of the phenotype modelling. Modelling was limited to a cross-sectional analysis, slightly underestimated the frequency of ankylosis and did not correct for some known influences of radiographic severity such as medication use and occupation. Robustness of the model to the effects of confounders was assessed by local sensitivity analysis.

As the first comprehensive study of the relationship between genotype and radiographic severity, this study confirms that genetic factors make a significant contribution. Involvement of loci highlighted by this study should be confirmed by larger genetic association studies and functional profiling of the new candidate genes in AS cases. Once identified, genetic determinants of radiographic severity would make attractive targets for therapeutic intervention and prediction of prognosis.

DECLARATION BY AUTHOR

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

PUBLICATIONS DURING CANDIDATURE

No Publications

PUBLICATIONS INCLUDED IN THIS THESIS

No publications included

CONTRIBUTIONS BY OTHERS TO THE THESIS

Chapter 2

Data collection was performed by members of Australo-Anglo-American Spondyloarthritis Consortium (TASC), including Diamantina Institute staff L Bradbury, K Hollis and J McFarlane. Radiographs were scored by Steven Truong, MA Stone, M Ward, TJ Learch, M Brown, L Bradbury and M Weisman. Steven Truong performed all statistical analyses, with guidance from P Leo, M Brown, N Warrington, D Evans and K Lê Cao. The analyses were conceived by M Brown and S Truong.

Chapter 3

Genotyping was peformed by J. Hadler, K. Cremin, K. Pryce and J. Harris. Quality control for Illumina Immunochip samples was performed by Adrian Cortes. Quality control of other samples and all genotype imputation was performed by Steven Truong, using protocols developed by P Leo, M Marshall and G Hemani. Statistical analyses, including GWAS, GCTA and allelic risk score analysis were performed by Steven Truong, with guidance from P Leo, D Evans and M Brown. The analyses were conceived by M Brown. R code used to create the heatmaps, Manhattan plots and Q-Q plots in this thesis were modified from code written by A Cortes or S Turner.

STATEMENT OF PARTS OF THE THESIS SUBMITTED TO QUALIFY FOR THE AWARD OF ANOTHER DEGREE

None

ACKNOWLEDGEMENTS

This work was only possible due to extensive collaboration. Data collection occurred in four nations and the analysis spanned multiple domains of science. I was only able to complete this complex project with assistance from experts in each domain, and would have been completely unprepared for it without decades of enthusiastic teaching and encouragement from primary school until my BSc.

I'd like to thank everyone at DI, for showing me that curiosity is not unusual in adults. I am grateful for my DI clinician-researcher scholarship and my UQ scholarship which provided enough to avoid the impoverished yet noble student life I enjoyed ten years ago.

I greatly appreciate the work and faith put into my project by my supervisors Matt and Paul. I value their guidance, and thank them for supervising me despite my lack of experience in research. Matt, I appreciate the introduction into the world of collaboration and Rheumatology research. Paul, thanks for teaching me how to deal with the frustrations of research - with patience and good humour.

Each week I also learnt about AS from the seasoned clinicians Phil, Linda and Kelly. Thanks also to Ranjeny and Helen for teaching me about early RA, and for showing me the role of the clinician-scientist.

Thanks Adrian for handling all my little questions, and for helping me turn overnight tasks into 3 minute ones. Thanks to the entire bioinformatics team - Paul, Mhairi, Adrian, Phil and Pamela, for help at every turn. Thanks to the Evans team – David, Nicole and John, for answering my most difficult questions and for patiently pointing me in the right direction when my statistics or statistical genetics veered into dangerous territory. To the rest of the Brown lab, thanks for the advice, company and enthusiasm.

Finally I must acknowledge the individuals who agreed to participate in our database, in the knowledge that they would be unlikely to receive any personal benefit. During this project I learnt that the best aspect of medical research, the privilege of researching *Homo sapiens sapiens*, is also the most challenging in a logistical sense.

Keywords

Ankylosing spondylitis, severity, bone formation, radiograph, x-ray, genetics, humans, Rheumatology

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 110322, Rheumatology and Arthritis, 60%

ANZSRC code: 110311, Medical Genetics, 40%

Fields of Research (FoR) Classification

ANZSRC code: 1103, Clinical Sciences 100%

Contents

LIST OF FIGURES

LIST OF TABLES

List of Commonly Used Abbreviations

AIC	Akaike Information Criterion
AS	Ankylosing Spondylitis
ASAS	Assessment of SpondyloArthritis international Society
BASDAI	Bath Ankylosing Spondylitis Activity Index
BASFI	the Bath Ankylosing Spondylitis Functional Index
BASRI	the Bath Ankylosing Spondylitis Radiology Index
BMD	Bone Mineral Density
BMP	Bone Morphogenic Protein
CIL	Corner Inflammatory Lesion
cmSASSS	contracted modified Stoke Ankylosing Spondylitis Severity Score
CNV	Copy Number Variant
D'	Linkage Disequilibrium coefficient
GCTA	Genome wide Complex Trait Analysis
GWAS	Genome Wide Association Study
HLA	Human Leukocyte Antigen
IBD	Inflammatory Bowel Disease
ICC	Intra-Class Coefficient
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
MHC	Major Histocompatibility Complex
MMP	Matrix Metallo Proteinase
MR	Magnetic Resonance
mSASSS	modified Stoke Ankylosing Spondylitis Severity Score
NB	Negative Binomial
NSAID	Non-Steroidal Anti-Inflammatory Drug
OR	Odds Ratio
PC	Principal Component
QTL	Quantitative Trait Loci
SNP	Single Nucleotide Polymorphism
SPARCC	The Spondyloarthritis Research Consortium of Canada
TASC	The Australo-Anglo-American Spondylitis Consortium
TNF	Tumour Necrosis Factor
UQDI	The University of Queensland Diamantina Institute
VAS	Visual Analogue Score
WTCCC	The Wellcome Trust Case Control Consortium
ZINB	Zero Inflated Negative Binomial

Thesis Introduction

Ankylosing Spondylitis is a highly heritable condition characterised by excessive bone formation of the spine and pelvis. A small number of non-genetic factors have been identified that explain less than one third of the variation in radiographic severity. Numerous studies have investigated genetic contributors to susceptibility. Few studies have compared genotype with radiographic severity and none have scanned the whole genome.

This study investigates the contribution of non-genetic and genetic factors to this phenotype with the aim of identifying causal variants that could be used to predict prognosis or as therapeutic targets.

1. Literature Review

This literature review will discuss the nature of radiographic signs of Ankylosing Spondylitis (AS). It will cover AS pathology related to radiographic changes, predictors of radiographic severity and the most appropriate method of assessing radiographic severity in AS. Identification of limitations of current knowledge will aid in focussing future research into AS severity.

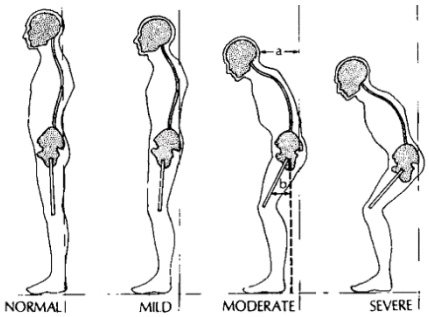
AS is a chronic form of autoinflammatory arthritis which affects approximately 0.5-0.86% of Europeans, and is two to three times more common in males (1, 2). The predominant symptoms of axial pain, stiffness and fatigue usually develop in the late teens or early twenties. These symptoms result from inflammation and bone formation of the spine, which is typically first evident in the sacroiliac joints. Entheseal inflammation occurs in the axial spine and many extra-axial locations, including the hip and knee joints, the plantar fascia and the insertion of the Achilles tendon. Other sites of inflammation can include: the anterior uvea, the intestines, the proximal aorta and the lungs. One third of sufferers will progress to complete bony fusion of the spine within thirty years of symptom onset, which confers significant disability. Recent Genome Wide Association Studies (GWAS) have identified genetic loci that contribute to AS susceptibility, but as yet few investigations have examined loci associated with AS severity. AS cohorts have been examined for non-genetic predictors of AS severity, but they explain only 23% of phenotypic variation in severity (3). Clinicians therefore have very limited ability to estimate prognosis in this commonly disabling condition.

The location of many prognostic factors is expected to be within the genome. Twin studies demonstrate more than 90% of AS susceptibility is explained by genetic variation (4). Pedigree studies of AS severity have estimated the heritability, or variation due to genetic factors, of a functional score, Bath Ankylosing Spondylitis Functional Index (BASFI), a symptom score, the Bath Ankylosing Spondylitis Activity Index (BASDAI), and score of radiographic severity, the Bath Ankylosing Spondylitis Radiology Index (BASRI). BASFI has an estimated heritability of 0.76, BASDAI 0.49, and BASRI 0.62 (5, 6). An alternate model of AS severity would be that while susceptibility is genetic, the progression of the

disease depends upon non-genetic factors present at the time. The limited amount of variation in AS severity explained by non-genetic factors, and the existing estimates of BASFI and BASRI heritability make this model unlikely. Thus the influence of genetics on AS severity is likely to be substantial, mirroring the magnitude of the influence of genetics on AS susceptibility.

Functional deficit from AS is mostly due to pain, stiffness, limitation of joint movement and fatigue. Serial symptom scoring of AS sufferers demonstrates large temporal variation in the symptoms of pain and stiffness (7). Current treatment of AS relieves axial pain, stiffness and associated fatigue in most patients. Restricted range of movement, measured by metrology, is not completely relieved by pharmacologic therapy and is highly correlated with radiographic signs of ankylosis. It remains the main cause of disability that clinicians are unable to treat. As radiographic signs of ankylosis are progressive and irreversible, they provide a replicable phenotype which can be objectively measured and is proportionate to disability.

Figure 1.1. Progression of spinal disease in AS (8)



The Pathology of Progression in AS

The pathological characteristics of AS and their progression have been investigated by Magnetic Resonance (MR) studies and histology. The sequence is believed to start with autoinflammatory osteitis, progresses to erosion of cartilage and bone, and is replaced by loose fibrous repair tissue which eventually ossifies (9). If this occurs at a vertebral corner a spur-like bony protrusion, termed a syndesmophyte, may form. Histological examination at autopsy has demonstrated that squared and sclerosed vertebrae on radiography are

sites of osteitis, infiltrated with neutrophils, lymphocytes and monocytes (10). Some areas of vertebral cortex were replaced by granulation tissue. Osteoclasts and osteoblasts were present, and woven bone was seen to be replaced by lamellar bone. Biopsy of tender entheses and ligaments from AS patients also has demonstrated inflammatory infiltrate. Entheseal erosions were present, and appeared to be repaired with woven bone without cartilage formation, then presumably was later replaced by lamellar bone forming a bony prominence at the site of former erosion (11). A similar process was observed at anterior vertebral corners, with the additional feature of endochondral ossification of the vertebral end plate. Some syndesmophytes were noted to have bases of lamellar bone and apices of woven bone. Given that lamellar bone usually replaces reactive woven bone after some time, this indicates intermittent bone formation. Fewer histological signs of inflammation were present at sites of new bone formation. These findings have been supported by MR studies of osteitis, which demonstrate syndesmophyte formation is more common in vertebral corners with observed osteitis at baseline (12). It is also more common at corners where inflammation has resolved, as evidenced by a fatty lesion or resolution of osteitis (13). These studies have also implied that syndesmophytes are more likely to occur at sites of advanced osteitis, called ‘type B’ corner inflammatory lesions (CIL), rather than acute ‘type A’ CIL, and where osteitis persists for the entire duration of the study. One study found a persistent CIL to be six times more likely, and a resolved CIL twice as likely to progress to a syndesmophyte than a vertebral corner with no CIL and no fatty lesion (13). This was in the context of an overall syndesmophyte formation rate of 3.7%, which is consistent with previously observed rates. Notably 40% of syndesmophytes in this two year study arose from corners with no observed inflammatory or fatty lesion (14). The lack of immediate bone formation at inflamed vertebral corners can be explained by the ‘TNF brake hypothesis’, where active inflammation and osteoclast stimulation temporarily inhibit bone formation (15). This may explain why an MR study comparing the syndesmophyte predictive value of two osteitis based scores did not find any benefit of including erosion information (16).

An alternative explanation for the lack of observed coexisting bone formation and inflammation would be that erosions do not progress to sites of bone formation. Lories et al. tested this theory in a mouse model by observing whether pathological bone formation

occurs when inflammation is treated with Tumour Necrosis Factor α (TNF) inhibitors and dexamethasone. Pathological bone formation from existing bone, termed osteoproliferation, was not affected (17, 18). The authors also suggested that a lack of inhibition of bone formation by glucocorticoids is a sign that osteoproliferative signals in AS differ from physiological bone remodelling signals, which are inhibited by glucocorticoids. Supporting the uncoupled theory, a subsequent mouse trial found that blockade of the osteoclastogenic signals TNF and RANKL did not affect bone formation, and sites of osteoclast infiltration were seen to progress to bony spurs. Sites where osteoclast infiltration breached the cortex and formed an erosion were observed to not progress to bony spurs (19). This mouse model develops an ankylosing enthesopathy focused around the ankle, and it is known that trauma (such as fighting between male mice) is involved its pathogenesis. Unlike most human AS cases, the mice involved do not carry *HLA-B27*, and do not develop axial arthritis. Thus the relevance of these mouse model findings to human AS is uncertain. Under the assumption that in this model bony spur formation occurred despite thorough suppression of all inflammation, the authors proposed an alternate model. This model treats inflammation and ankylosis as uncoupled processes, and states that biomechanical stress or damage triggers a local reaction that may result in either osteogenesis or inflammatory cell infiltration. At this point there is no human radiographic data to support this model, and the strong link between CIL and syndesmophyte formation suggests a coupled process. Changes to bone including osteitis, vertebral erosion, sclerosis and squaring have been implicated in the pathogenesis of syndesmophyte formation, but the exact pathological process that leads to their formation is not clear.

The only known mechanism of bone resorption is by osteoclasts. Osteoclasts are overactive in AS, demonstrated by a greatly increased incidence of trabecular osteoporosis. Osteoblast precursors, lymphocytes, Natural Killer (NK) cells and synovial cells all express RANKL, a signal which promotes osteoclast differentiation from monocyte lineage precursors and stimulates resorption (20). This process is directly inhibited by osteoprotegrin. Another inhibitor of osteoclast differentiation is IL-12 (21), which is associated with Single Nucleotide Polymorphisms (SNPs) in AS (22). RANKL production is induced by TNF, IL-1, IL-6 and IL17/23 with IL-6 also exerting a direct inhibitory effect on

osteoclasts. Apart from IL-17, the receptors for these cytokines have been associated with SNPs in AS. TNF also acts synergistically with RANKL on osteoclast precursors to promote osteoclast differentiation. Osteoclast formation in response to TNF is 2.5 fold higher in *HLA-B27* positive monocytes, an allele present in most AS patients (23). IL-12 and IL-23 inhibit osteoclast formation by inducing production of Granulocyte Macrophage Colony Stimulating Factor (GMCSF) which promotes mononuclear cell differentiation into dendritic cells rather than osteoclasts (24). Thus the spondyloarthropathies, which are mediated by a Th17 response involving IL-12, IL-23, IL-17, IL-1, TNF and IL-6, produce a mixture of inhibitory and stimulatory osteoclast signals. Bone resorption is regulated by complex overlapping cytokine pathways.

Bone formation is also regulated by the same pathways. Some cytokines involved in the Th17 response, such as IL-23, also stimulate bone formation (25). TNF and IL-6 inhibit bone formation by upregulating DKK-*1* and sclerostin, negative regulators of the 'wingless' (Wnt) pathway, a pathway central to osteoblast differentiation and function (21). Blocking these inhibitory signals promotes bone formation, as evidenced by increased osteophyte formation, by endochrondral ossification, upon combined DKK-1 and TNF blockade (26). Wnt dysregulation in AS has been demonstrated by the observation that DKK-1 and sclerostin levels are lower in subjects who develop syndesmophytes over a two year period. DKK-1 levels in the AS subjects were not correlated with CRP, and were noted to be stable over three measurements spanning two years (27). Kwon et al. found that DKK-1 levels in AS cases are almost half that in healthy controls, and are unresponsive to three months of TNF blockade (28). This suggests DKK-1 levels are independent of acute inflammation, and may be instead influenced by other signalling mechanisms, host factors such as genetics or may display a delayed rise after inflammation. A similar study by Daoussis et al. found the opposite – with DKK-1 levels higher in AS cases than controls, and DKK-1 levels higher in the TNF treated group (29). This study also found DKK-1 levels in RA subjects to be lower than in controls, which is highly inconsistent with previous studies of DKK-1 levels, and the fact that DKK-1 blockade induces sacroiliac fusion in TNF transgenic mice (30). Bone Morphogenic Proteins (BMP), members of the Transforming Growth Factor β family, stimulate physiological chondrogenesis and bone formation. BMP dysregulation is evident in mouse models (17, 18), and one study of human AS cases

suggests BMP is elevated relative to controls in AS subjects with spinal fusion but not AS subjects without fusion (31). Matrix Metalloproteinases (MMPs) break down extracellular matrix during remodelling, and may be a marker of AS activity. Elevated levels of MMP3 have been linked to radiographic progression over two years (32). A follow up study did not find MMP levels elevated in the bone or synovium of AS cases, leaving their role in AS undefined (33). Dysregulated bone formation and remodelling pathways are implicated in syndesmophyte formation, and appear to be influenced by inherent host factors and inflammatory pathways.

Numerous inflammatory pathways have been implicated in AS pathogenesis. Studies from AS cases and mouse models have strongly implicated Th17 cytokines IL-17 and IL-23. They are produced at sites of inflammation, mostly by innate immune system cells including mast cells and γ/δ T cells (25, 34). This response is also seen in immune response to Chlamydial infections, a trigger of reactive arthritis. IL-23 triggers inflammation of entheses and the proximal aorta in mouse models, and promotes IL-22 expression which in turn upregulates Wnt and BMP (25). Overactivity in these pathways is deliberately introduced into mouse models, while its origin in humans is believed to be mostly inherited. Genetic studies have found association between AS susceptibility and SNPs in receptors for IL-23, IL-6, IL-12, IL-27 and prostaglandin E2 (35, 36). TNF, IL-1, IL-6, IL-17 and IL-23 are considered shared pathways of multiple physiological inflammatory responses (34). Elevated IL-1 and IL-6 levels are seen in gout, while rheumatoid arthritis is associated with increased production of TNF, IL-6, IL-17 and IL-23. Prostaglandin E2 is produced at sites of inflammation, and promotes osteoblast differentiation via Wnt activation. Non Steroidal Anti-Inflammatory Drugs (NSAIDs) limit radiographic progression in AS, an action attributed to decreased prostaglandin production (37-39). Other loci implicated in AS pathogenesis include those involved in lymphocyte differentiation and function, peptide handling and extra-cellular matrix interaction. A GWAS of Han Chinese cases associated *HAPLN1*, an extra-cellular matrix linker protein associated with osteophyte formation in postmenopausal women (40). This study also associated *EDIL3*, also known as *DEL1*, which inhibits bone resorption in a periodontitis mouse model by downregulating IL-17 and subsequently RANKL (41). No association with either *HAPLN1* or *EDIL3* was found in two GWAS of AS susceptibility in Europeans

and another study including 2 998 East Asian cases and 5 547 East Asian controls (22, 35, 36). Other immunological genes associated with AS susceptibility include the aminopeptidases *ERAP1*, *ERAP2*, *LNPEP* and *NPEPPS* (22). They cleave antigens before presentation in HLA I and have been hypothesised to increase AS risk by endoplasmic reticulum stress or aberrant antigen presentation to T lymphocytes and Natural Killer cells. AS pathogenesis involves a prominent Th17 response as well as production of cytokines common to multiple autoimmune diseases. This is reflected by the susceptibility loci identified on GWAS.

Identifying Genetic Variation

GWAS is an efficient method of identifying cellular mechanisms responsible for disease. GWAS measure hundreds of thousands of SNPs across the genome, and compare this to a known reference sequence or haplotype map to infer the sequence of nearby variable regions. Measured SNPs display high Linkage Disequilibrium (LD) to nearby loci. Distant loci are less likely to be inherited together, as the chance of recombination is proportionate to genetic distance. A SNP is considered associated with a disease if there is a significant difference in allele frequency between cases and controls, or between subjects with different levels of a quantitative phenotype. Variation in a SNP or in the surrounding region with high LD may contribute to pathogenesis. These regions are examined for coding or regulatory sequences related to the affliction. This method allows examination of most of the European genome while genotyping fewer than a million SNPs, despite over 38 million human SNPs being identified (42). This high throughput approach is cheaper and faster than an equivalent proteomics based approach, where a small number of candidate pathways are identified through slow, costly and laborious examination of case and mouse model tissue. The search for diseased mechanisms in human cases and mouse models usually requires knowledge of where to look, what to measure, and depends on mouse models that accurately reflect the disease in humans. GWAS does not investigate ‘rare variants’, it excludes SNPs with a Minor Allele Frequency (MAF) <1%. The total contribution of rare variants to polygenic disease is controversial. Rare variants of genes *CARD9*, *IL23R*, *LNPEP* and *TYK2* have been found to contribute to AS

susceptibility (22). Most GWAS testing does not include specific tests for sporadic mutations, epigenetic changes or genomic structural features such as Copy Number Variants (CNV) and insertions and deletions. Despite this, GWAS tag most CNV variation, meaning that most disease associated CNV should appear as disease associated SNPs on GWAS (43). GWAS is an efficient method of identifying pathological processes, but has limitations and has only identified 24.4% of the heritability of AS susceptibility, compared to an estimated total heritability of >90% from twin studies (4, 22).

Heritability estimates from twin studies and pedigree data are considered robust calculations. The heritability with no identified source is termed 'missing heritability'. While some 'missing heritability' may exist undiscovered in rare variants, epigenetics and other known non-SNP sources of variation, an effort is being made to quantify the contribution of common variants with small effects (44). It has been suggested that much of the missing heritability is spread across hundreds or thousands of common SNPs, each with an effect size too small to pass the stringent genome wide significance tests used to limit false positives in GWAS (45). Advocates of this hypothesis have developed a method of estimating heritability from additive genetic variants (narrow sense heritability) using SNP data, called Genome Wide Complex Trait Analysis (GCTA). This method, which estimates relatedness by allele sharing, has less error than similar methods that estimate relatedness using detected Identity By Descent segments (46). Application of GCTA to Parkinson's Disease attributed 27% of narrow sense heritability to SNPs, a figure much closer to estimates from family studies than the 3-5% from GWAS hits (47). This analysis allows researchers to quantify variation responsible for heritability without specific knowledge of which SNPs are contributing. Comparisons of narrow sense heritability estimated by GCTA and heritability calculated from familial studies can partly answer the problem of missing heritability, confirming that common SNPs are a major source of phenotypic variation.

Measuring the Radiographic Severity of AS

Symptoms and radiographic changes in AS are predominantly related to the axial skeleton. Initial symptoms of buttock and lower lumbar pain coincide with sacroilitis, initially visible on MR imaging, which is followed over years by sclerosis, erosion, osteoproliferation and joint fusion on plain radiography. These contrasting signs of erosion and bone formation reflect the complex signals that promote bone resorption while also promoting bone formation, and provide specific criteria for radiologic diagnosis. For over twenty-five years, definite AS has been diagnosed by the modified New York Criteria, which require typical symptoms or changes in metrology plus definite sacroilitis bilaterally or severe changes unilaterally (48).

Early radiographic changes seen at vertebral corners include erosions, sclerosis and squaring, with syndesmophytes and bony bridges between vertebrae occurring later. The modified Stoke Ankylosing Spondylitis Spinal Score (mSASSS) has been validated as a reproducible severity score which is strongly correlated with metrology, progression and functional scores (14, 49-51). mSASSS has been endorsed as the most appropriate scoring method for trials by the Assessment in Ankylosing Spondylitis working group (52). It scores one point for any erosion, sclerosis or squaring at twenty-four anterior vertebral corners in the cervical and lumbar spine, two points for a syndesmophyte and three for ankylosis. The omission of posterior corners is valid – as anterior vertebral fusion is more strongly correlated with BASFI than posterior fusion, and posterior corners are often not clearly visible on plain radiograph (50). Facet joint ankylosis is a prominent, function limiting feature of AS, but their use in scoring systems has been limited as lateral views are not the best angle for assessment (51, 53). Studies have consistently found strong correlation between mSASSS and outcomes including BASFI, AS metrology, and hip damage measured by the BASRI hip score, and the BASRI spine score, a four level severity score (49, 50, 54). Due to its reproducibility and correlation with functional outcomes, mSASSS is the preferred method of scoring spinal changes in AS trials (52).

Figure 1.2. Radiographic disease progression in a single patient. Zero syndesmophytes progress to complete ankylosis.

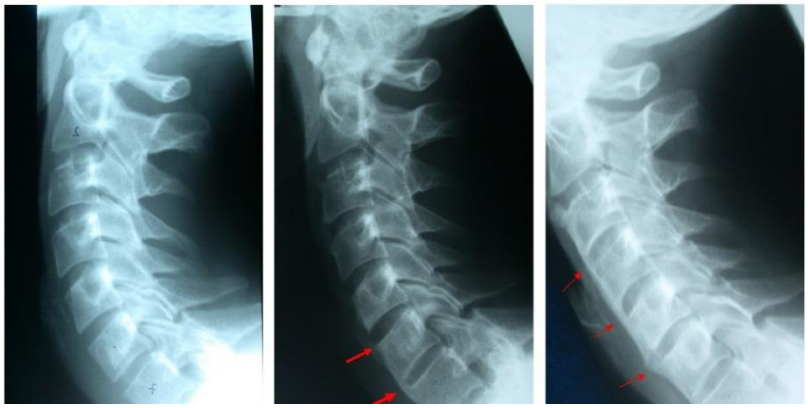
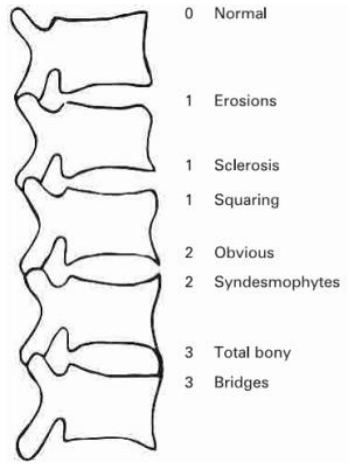


Figure 1.3. The mSASSS scoring system (48).



An ideal score of cumulative AS severity should be based on features that directly contribute to functional impairment. Presence of osteoproliferation, indicated by syndesmophytes and bony fusion, directly decrease spinal range of movement and therefore function. Removal of ambiguous syndesmophytes from mSASSS scoring by not counting any bony protrusions growing greater than 45° from the anterior plane of the vertebra improved the ability of mSASSS to predict progression (14). Numerous researchers have suggested removing scores for cervical squaring as it is not consistent between scorers, and occurs in <1% of vertebral corners. They have also criticised the scoring of erosions as erosions may represent a pathologic process separate to or distant from syndesmophyte formation, and have minimal impact on mSASSS as they occur in <5% of changes over two years (14, 55, 56). The only studies that examine the future of

erosions, sclerosis and squaring (all from the OASIS cohort) suggest that erosions and sclerosis but not squaring have an odds ratio of 2 for progressing to syndesmophytes over a twelve year period (57, 58). Despite these criticisms, the reliability of total mSASSS has been repeatedly demonstrated, with high intra- and inter-observer correlations for cross-sectional data (49, 51, 54). Longitudinal data measured as change in score between films demonstrate higher inter-observer correlation (55, 59). To minimise error, it has been suggested that the use of multiple scorers minimises bias from a single scorer (51). To improve sensitivity to change, films from all timepoints from the same subject should be scored at the same time (60). While the mSASSS is a reproducible score predictive of future progression, scoring of non-syndesmophyte elements may detract from its role as a measure and predictor of ankylosis.

Many of these studies aiming to validate mSASSS have also provided data into the natural history of AS. It has been observed that 27.9% of subjects progress to complete spinal fusion after thirty years, and 42.6% after forty years (61). Progression has been noted to be linear, irrespective of disease duration or age, and the proportion of patients with lumbar, cervical and hip involvement has been demonstrated to increase linearly over time (62). The average rate of progression has repeatedly been found to be approximately one mSASSS point per year, with 33% of subjects progressing within two years of observation, 48% within four years and 82% after twelve years in the OASIS cohort. As scores of one are uncommon, one mSASSS point per year is equivalent to one new syndesmophyte every two years. While many subjects did not progress, a subset progress rapidly, with 22% of subjects in OASIS cohort and 38% of a German cohort progressing at least twice as fast as average (63-65). As progression is almost universal, mSASSS is proportionate to disease duration. The recommended film interval is every two years to provide sensitivity to change and adequate time for progression. Two cohorts scored with BASRI spine have demonstrated a predominance of syndesmophytes in the lumbar compared to cervical spine in the first twenty years of symptoms (61, 62). The natural history of AS is slow, linear progression for most patients.

Determinants of Radiographic Severity

The *HLA-B27* allele is the largest identified genetic contributor to AS susceptibility, and has repeatedly been investigated for a link to AS radiographic severity. Exactly how it predisposes individuals to AS, or why the majority of B27 positive individuals do not develop AS is not clear. Theories of its mechanism of action in AS pathogenesis include that it may present a particular arthritogenic peptide to induce autoimmunity, that its subunits may be targeted by antigen presenting cells, or that it may contribute to endoplasmic reticulum stress and trigger IL-23 production (66). *HLA-B27* variants, and the presence of B27 heterozygosity or homozygosity has not consistently been linked to AS severity. Boonen et al. found no association between HLA-B27 positivity and syndesmophyte presence or spinal fusion in 498 subjects (67). Ward et al. found no association between *HLA-B27* and BASRI spine score in a cross sectional study of 398 subjects. Studies of Finns demonstrated no relationship of *HLA-B27* positive homozygotes with BASFI or BASDAI, but confirmed increased risk of AS susceptibility compared to *HLA-B27* heterozygotes (68). A cross sectional Turkish study found of more severe spinal disease in *HLA-B27* positive males compared to negative males on BASRI, but only in the lumbar spine (69). A recent longitudinal Korean study of 127 homozygous positive and 241 heterozygous cases found no influence on mSASSS progression, BASFI or BASDAI, after a mean follow up of 2.86 years and correction for clinically relevant covariates (70). The longest longitudinal data is from the OASIS cohort of 186 subjects, with an average follow up point of 7.9 years. *HLA-B27* positive men were found to progress faster than *HLA-B27* negative men at rates of 1.18 and 0.69 mSASSS units per year respectively upon univariate regression (65). The authors provided no values corrected for the influence of the other covariates. The lack of a corrected progression rate, the fact that it only occurs in males, and the absence of this finding on multiple previous cross sectional and one corrected longitudinal study mean that the body of evidence does not support a major role for HLA-B27 in the variation of AS radiographic severity. While HLA-B27 is clearly involved in AS susceptibility, it remains unclear whether it is associated with the radiographic severity of established AS.

No genetic association studies into determinants of AS radiographic severity have been performed on a genome wide scale. Candidate gene studies have attempted to associate this quantitative trait with genetic loci, termed QTLs (Quantitative Trait Loci). A candidate gene study of 398 subjects compared the *HLA-DRB1, -DQA1, -DQB1* and *-DPB1* alleles to BASRI spine scores by multivariate logistic regression. This cross sectional study found *HLA-DRB1*0801* to be protective, and *DQB1*0603* and *DQB1*0401* to predict severe disease. These alleles were present in 6%, 8.3% and 9.5% of subjects respectively (71). Another candidate gene study that examined 384 SNPs from immunological and bone related loci found significant associations with: a β adrenergic receptor *ADRB1*, the osteoblast stimulator *NELL1*, and SNPs near *HLA-B*, *-DRB1* and *-DQA1*, whereas *HLA-B27* did not display a significant association (72). The third candidate gene study of radiographic severity compared thirteen SNPs tagging five genes, with mSASSS scores from 241 Canadian cases, with a mean follow up of 2.4 years. This study demonstrated an association with *LMP2* and baseline mSASSS, after correction for gender and disease duration. This association did not exist with mSASSS progression, in fact only baseline mSASSS had a significant predictive value (73). *LMP2* is also known as *PSMB9*, and is a proteasome subunit involved in peptide handling and antigen processing. The significance of these candidate gene study results is difficult to interpret given that none of these loci have been identified as susceptibility loci in genetic studies of AS susceptibility. A recent Immunochip study with good HLA coverage, found an association with *HLA-A*0201* and *HLA-B27* and AS susceptibility, but not with *HLA-DRB1* or the other *HLA-B* alleles found in the candidate gene study of radiographic severity (22). The non-HLA QTLs found by these studies have not been associated by AS susceptibility GWAS or Immunochip either. This implies that either these reported QTLs were false negatives on the susceptibility GWAS, false positives as QTLs, or influence severity but not susceptibility. The latter possibility is unlikely - QTLs are expected to be related to disease pathogenesis and so should also contribute to disease susceptibility. A QTL that does not contribute to susceptibility would have to affect a disease related pathway many steps away from the initiating pathological events. To date, GWAS of other autoimmune diseases have not definitely identified such an entity, with QTLs being previously identified susceptibility loci, as in an asthma severity GWAS (74), or no QTLs being found at all, as in Multiple Sclerosis (75, 76). These inconsistent findings, the patchy haplotype coverage of

candidate gene studies and a lack of correction for population stratification and multiple testing make the relevance of these loci uncertain (77). Candidate gene studies have pointed to a number of QTLs which should be investigated as determinants of severity by a more comprehensive genetic study.

The most visible genotype that predicts severity in AS is the male gender. Many cross-sectional and longitudinal studies, after correcting for relevant covariates, have found the male gender to be associated with more severe radiographic disease and faster progression (3, 62, 65, 69, 71, 78). The rate of progression (before correction) was found to be significantly higher in males, at 1.11 mSASSS units per year compared to 0.69 in females (65). Conversely two longitudinal studies and one cross-sectional study analysing predictors of progression found no association with gender after correcting for other factors, despite the largest having 334 subjects, including 22% females (63, 71, 79). The reason why such a strong effect was found in most but not all studies is not clear. Gender based differences in the immune system are well described, and are attributed to regulatory effects of oestrogen on leukocytes and X inactivation. Notably, decreased IL-1 and IL-6 production occurs in the presence of oestrogen (80). Many of the genetic causes of immunological gender differences, such as dosage compensation of X linked genes may also apply to AS related loci. Oestrogens are also known to inhibit osteoclast function and to stimulate bone formation by stimulating Wnt in osteoblasts, while androgens stimulate bone formation in men (81). This anabolic effect of androgens in males may add to other osteogenic signals in AS, favouring osteoproliferation. Gender based patterns of severity exist in AS. Presence of lumbar syndesmophytes is less common in females, but peripheral arthritis is more common (61). Some studies have demonstrated a predominance of cervical spine disease in females, while others have found no difference (61). Females have slightly higher BASFI for equivalent BASRI scores compared to males, which has been attributed to the higher incidence of peripheral arthritis (78). Gender is strongly associated with AS radiographic severity, which is presumably due to recognised effects on the immune system and bone.

Cigarette smoking has also been strongly linked to dysregulation of the immune system and bone metabolism. Current smoking has repeatedly been associated with mSASSS progression, in a dose dependent fashion (71, 79). Analysis of subjects with AS within the GESPIC cohort found that current heavy smokers (>10 cigarettes per day) progressed significantly faster than others, increasing their mSASSS on average 3.1 units over two years compared to 0.57 units over two years for moderate and non-smokers (82). In another early spondyloarthritis cohort, which contained subjects with and without definite AS, higher mSASSS scores and spinal inflammation on MR were more common in current smokers (83). Smoking has long been recognised as a proinflammatory factor, and it also causes elevated IL-1, IL-6 and GMCSF, all of which inhibit osteoclast formation (21). Clinical evidence and knowledge of immunological and bone metabolism pathways support the notion that smoking is a determinant of radiographic severity in AS.

Prostaglandin signalling is an immunological pathway involved in bone metabolism. NSAIDs have been used to treat inflammation and pain in AS, and have been demonstrated to limit radiographic progression over a two year period (37-39). Recently published data from two longitudinal cohorts found no impact of NSAID intake, as measured by the ASAS NSAID index, on radiographic progression (65, 79). The three studies that demonstrated an effect of NSAIDS found a large effect size, with the odds ratio of risk of progression being 0.15-0.36. The first study had a borderline p value of 0.045 (37), the first and second studies only found an effect in subjects with raised inflammatory markers (37, 38), and the third did not match cohorts for baseline mSASSS (39), thus leaving the greatest predictor of mSASSS progression (65, 79), as a confounding factor. While NSAID induced inhibition of radiographic progression is consistent with theory, further longitudinal studies should be conducted to confirm the effect size of this medication.

The other major class of medication used in AS is TNF inhibitors. If TNF induced inflammation is an immediate trigger of osteoproliferation, TNF inhibition should impair radiographic progression. Randomised Control Trials (RCTs) of etanercept and golimumab and case control studies of adalimumab and infliximab found no effect on

mSASSS progression after two years (59, 84-86). Haroon et al. examined the effect of any TNF inhibitor use in a propensity score matched observational study, with a mean follow up of 2.9 years. TNF inhibitor use was found to be protective in multivariate logistic regression assessing risk of progression, with subjects commencing TNF inhibitors >10 years after symptom onset more likely to progress. Longitudinal analysis only revealed an effect in subjects followed up for more than 3.9 years (79). This study suggests that TNF may limit mSASSS progression, especially if started soon after symptom onset, and that the effect may only be evident after 3-4 years, explaining why the two year randomised control trials found no effect. The theory that only suppression of early inflammatory lesions with TNF inhibits bone formation is consistent with findings that advanced 'type B' CIL are more likely to progress to syndesmophytes than early 'type A' CIL (13). Thus if TNF inhibitor use prevents type A CIL from progressing to type B CIL, it may only inhibit bone formation if commenced early in the disease. TNF inhibition has been demonstrated to decrease erosions in rheumatoid arthritis, and to improve trabecular osteoporosis in AS, but its overall effect on the complex, interacting erosive and osteoproliferative pathways of AS is controversial (87-89). Given the conflicting clinical evidence and different theories regarding its effects on bone metabolism, further longitudinal and animal studies are needed to reveal the exact effects of TNF treatment.

Multiple clinical parameters have been repeatedly associated with mSASSS progression. After disease duration, presence of syndesmophytes at baseline appears to be the second best predictor of mSASSS progression (14, 63, 64, 79). This consistent finding, coupled with the average rate of mSASSS progression suggests that absence of syndesmophytes after sufficient disease duration is a protective prognostic marker. Severe hip disease is strongly correlated with BASRI spine score (3, 61, 62). This close relationship is presumed to be an extension of the AS disease process that affects the axial spine, although hip osteoarthritis is common and has a similar radiographic appearance. Hip disease in AS exhibits almost the same correlation with BASFI as mSASSS (50), making it an important functional outcome as well as a predictor of spinal disease.

Disease activity has been inconsistently linked with risk of syndesmophyte formation. ESR or CRP elevation indicate inflammation as part of an acute phase response. As discussed earlier the pathological relevance of acute inflammation to syndesmophyte formation is poorly understood. Bone formation and overactivity of the IL-17/23 pathways can occur without ESR or CRP elevation, and small clinical trials of IL-6 blockade in AS have not demonstrated any symptomatic benefit (90, 91). Nevertheless ESR and CRP may be markers of syndesmophyte formation, without being part of the causal pathway themselves. mSASSS progression has been associated with baseline inflammatory markers in a Canadian AS RCT and German AS RCT, and baseline and time averaged inflammatory markers in AS cases from the observational GESPIC cohort. No association with BASDAI was found in any of these studies (79, 85, 92). A recent update from the OASIS cohort, analysing longitudinal data found ESR, CRP, BASDAI and ASDAS (another a disease activity score) were associated with radiographic progression over the subsequent two to twelve years (93, 94). A report from this cohort at four years, found that BASDAI and CRP had no association with cross-sectional mSASSS univariate modelling, while ESR and ASDAS had p values of borderline significance (0.07 and 0.06 respectively) (64). An RCT on NSAID use found no effect of ASDAS, BASDAI or time averaged CRP on radiographic progression, but time averaged ESR was significant (38). Thus the association between measures of disease activity and radiographic progression, even from within the same cohort, are inconsistent. Further investigation is needed into these markers, along with whether inflammation precedes syndesmophyte formation.

Lifestyle factors may also influence radiographic severity. Physical activity and AS specific stretches are cornerstones of AS multidisciplinary management. They have been shown to improve symptom scores but not radiographic severity (3, 95). Conversely, occupations that require frequent bending, stretching, twisting or exposure to whole body vibration have been associated with higher BASFI and BASRI scores. Work environments involving sitting, stranding and kneeling did not demonstrate any association (96). A previous study that classified occupations as sedentary, active or manual, found no association with BASRI (3). Conversely a recent report (published only as an abstract to date) from the OASIS cohort found that blue-collar workers had higher rates of progression than white-collar workers (97). Level of education has recently been linked to mSASSS progression,

with individuals with university degrees having the lowest risk of progression (98). The authors have suggested that level of education is a proxy for socio-economic status, which may explain the association. However level of education may also be representative of job activity level, exercise, medication and stretching compliance and other healthcare related behaviour. Occupational and social factors that affect disease progression are important to identify as they may be modifiable.

Extra-spinal manifestations of AS, including anterior uveitis, psoriasis and peripheral arthritis have not consistently been linked to radiographic severity. A past history of uveitis has been associated with radiographic severity in three observational studies examining subjects from Bath (3, 62, 99). An effect persisted after multivariate analysis which corrected for gender, disease duration, and smoking status. Other cohorts do not demonstrate a correlation, even one cohort with a mean follow-up of 7.9 years (65, 67, 69, 72, 82). Comorbidity of AS and psoriasis is common, with 9.3% of AS sufferers reporting psoriasis (100). Follow up of the OASIS and two smaller cohorts has not demonstrated an impact of psoriasis on radiographic severity (65, 92, 101). An earlier study comparing matched controls to 91 AS subjects with psoriasis, and to 23 AS subjects with Inflammatory Bowel Disease (IBD) found no difference in BASRI spine scores (99). Two recent small studies have found peripheral arthritis (with or without psoriasis), affecting joints other than the spine or hips, to be predictive of slower spinal radiographic progression (69, 102). Given the closely related genetic risk factors for susceptibility of AS, psoriatic arthritis and IBD (103, 104), a subset of patients may have a mixed phenotype, involving features uncommon in AS but common in psoriatic or IBD associated arthritis, such as upper limb arthritis and an absence of spinal ankylosis. These patients might meet the modified New York criteria for AS but progress slowly. Presence of extra-spinal manifestations allows identification of subgroups with different pathology that may experience different degrees of ankylosis.

Patterns of Severity

Juvenile onset AS, defined as onset before age seventeen, has been linked to higher BASRI hip scores, a higher rate of hip replacement and lower BASRI spine scores (105). While the finding of lower lumbar spine BASRI in this subgroup was replicated in a Taiwanese cohort, worse BASRI hip and cervical spine scores were not (106). Two further studies found the association between juvenile onset and lower overall mSASSS scores (without reporting cervical and lumbar scores), but no significant association with BASRI hip score (107, 108). An increased prevalence of peripheral arthritis and dactylitis has been noted in juvenile onset AS (106, 107). In adult onset AS, multiple cross-sectional studies have found more severe disease in subjects with an older age of onset (71, 72). This has not been demonstrated in longitudinal studies dedicated to identifying risk factors for progression (14, 65, 79, 95). Factors associated with a late age of onset include absence of HLA-B27 and the male gender (68, 78). Patterns of AS severity appear to vary with age of onset, which may represent overlapping pathological mechanisms.

mSASSS modelling

Count data, including radiographic scores, often involve large proportions of low scores. Analysis of non-parametric count data requires the use of less commonly used modeling methods. Analysis of radiographic severity in AS has only been performed in a few cohorts, utilizing three distinct approaches. The first approach involves purely descriptive and observational statistics. Baraliakos et al. reported observed rates of syndesmophyte formation in a German cohort, and did not mention the distribution of the data (14). This study reported descriptive statistics (eg the mean) that do not require the data to be normally distributed. The second approach has been used to analyse the OASIS cohort of AS. Rates of change were reported, with no assumption of a normal distribution. Modeling was performed using Generalised Estimating Equations (GEE) (57). GEE can be used to analyse repeated measurements where there is correlation between observations (eg longitudinal data) but the distribution of correlation is unknown (109).

GEE require correlated response variables and so are not appropriate for analysis of cross-sectional data.

The third method used to analyse mSASSS was first used for this condition by Haroon *et al.* in a longitudinal study (79). Notably the cohort studied by Haroon *et al.* overlaps the cohort analysed in this MPhil project. Haroon *et al.* used a Zero Inflated Negative Binomial (ZINB) regression model for cross-sectional and longitudinal analysis of mSASSS. The authors justified the choice of model with the finding that change in mSASSS was over-dispersed and zero-inflated, with 55.4% of delta mSASSS values being zero. Parameters describing the goodness of fit of this model were not published.

Overdispersion, where the variance is higher than the mean of the observations, is a common problem when modeling count data (110, 111). If not corrected, overdispersion can cause model misspecification. Apparent overdispersion can result from model misspecification issues such as a missing covariate or interaction, outliers of the response variable and non-linear predictor effects. Real overdispersion may result from truly large variance, an excess of zeros, correlation between observations or clustering of observations (112, 113).

The Poisson distribution is equidispersed, where the variance equals the mean. In the negative binomial distribution the variance may exceed the mean. The Negative Binomial (NB) distribution is a mixture of the Poisson and gamma distributions (112, 113). It is appropriate for count data that consists of non-negative integers, in which the overdispersion is believed to occur in a gamma distribution. Gamma distributions include a wide range of shapes including a Gaussian distribution and left and right skewed bell shaped curves. If the overdispersion forms a recognizable distribution such as an inverse Gaussian distribution, modeling is typically performed by simulation (113). Otherwise “If overdispersion in the data takes no identifiable shape, most statisticians employ a negative binomial” (113)(Hilbe, p2).

Overdispersion in Poisson or NB models that persists after revision of the model terms and investigation into the previously mentioned causes over overdispersion is commonly caused by excessive counts of zero (zero inflation). The best method of addressing zero inflation is to use a zero inflated model (113). Zero inflated models are mixture models. They combine two models – one predicts the outcome of a zero vs a non-zero value (binomial model), and the other predicts the outcome for a range of non-zero ('count') values.

Assessment of model fit for Poisson and NB models differs from that of other models types. Notably these models allow a large amount of variance, allowing a greater difference between fitted and observed values than is present in other model types. The residuals must be considered in proportion to the variance, and so are scaled by the square root of the variance as Pearson residuals. Common methods of model validation for negative binomial models include observed vs fitted plots, plots of predictors vs residuals, and residuals vs predicted plots (112). These should be assessed for homogeneity. Importantly, plots of residuals should not be expected to be normally distributed. "As we are now using techniques that do not need a normal distribution, a normal distribution of the residuals is no longer of concern" (110) (*Zuur et al.,* p86). Instead residuals may be assessed by comparing them with an envelope of bootstrapped confidence intervals (114, 115). If the confidence interval completely contains the residuals the model is likely adequate. This method assesses for incorrect model specification of error distribution, the influence of outliers and for missing explanatory variables.

Summary

Individuals with AS experience symptoms across numerous areas of their body, and experience great variation in severity. More than a third develop complete spinal fusion, almost a third grow syndesmophytes twofold faster than average while many never develop spinal fusion. The pathophysiology that leads to osteoproliferation is only partly understood, leaving no definite explanation for features seen on MR and radiographic

studies. This leaves clinicians with very little ability to predict which patients will progress to spinal or hip ankylosis.

The slow rate of radiographic progression of AS and the relatively small size of AS cohorts has limited identification of determinants of severity. The established determinants of radiographic progression are existing syndesmophytes, genetics, disease duration, gender, smoking and continuous NSAID use. TNF inhibitor use may slow progression, while raised inflammatory markers and occupation-induced mechanical stress appear to be associated with mSASSS progression.

Previous research suggests the greatest determinant of AS severity is within the genome. No study has compared comprehensive genetic data with AS radiographic severity. Research in to genetic and non-genetic determinants of radiographic severity should provide valuable prognostic information and therapeutic targets that will improve the lives of AS patients.

2. Cross-Sectional Analysis of Radiographic Severity in AS

2.1 Introduction

Radiographic severity in AS varies from no spinal changes to complete ankylosis, and known factors explain a minority of this variation. Previous research has repeatedly identified a number of factors – symptom duration, gender, smoking status, continuous NSAID treatment, existing syndesmophytes, specific bone metabolism markers, and AS hip involvement to be predictive of radiographic severity. There are also select reports that TNF inhibitor use, occupation, AS symptom scores and inflammatory markers may predict radiographic severity.

This cross-sectional analysis aims to describe patterns of severity, and to assess the influence of recorded covariates on a cohort of over 1000 subjects. After removing the influence of these covariates, the data could be used as a phenotype for genetic association testing. With adequate power and replication, this could be used to create a prognostic model of AS radiographic severity, or to identify genetic and non-genetic areas for further research.

2.2 Methods

2.2.1 Radiographic and Clinical Data

Subject data were included from The University of Queensland Diamantina Institute (UQDI), and The Australo-Anglo-American Spondylitis Consortium (TASC). All subjects had definite AS by the modified New York criteria (48). UQDI films were scored with mSASSS by either Steven Truong (ST), or Linda Bradbury (LB) with Matthew Brown (MAB). TASC films were scored once by LB, MAB, Thomas J Learch (TJL), Millicent Stone (MS), Michael Ward (MMW) or Michael Weisman (MHW). Written informed consent was obtained from each subject and the data collection was approved by the ethics committee at each centre.

mSASSS scoring by ST was compared to other TASC scorers using twenty-two TASC reference films and thirty-nine UQDI films, each from a different subject. ST was blinded to previous scores at the time of scoring. When scoring multiple films from the same subject, ST scored films chronologically and was not blinded for their date. Blinding for date is technically difficult (they are date stamped) and was not considered necessary by mSASSS validation studies (49, 51, 60). The most recent mSASSS for each subject was used, unless it indicated complete cervical or lumbar ankylosis. In these cases, where possible the score from the film preceding complete ankylosis was used to avoid a 'ceiling effect'.

In the case of mSASSS a 'ceiling effect' can exist because scores are right censored in a biological sense. mSASSS scores cannot be greater than 72 but clinical severity continues to worsen after mSASSS reaches 72. Once a subject has an mSASSS of 72, even if the disease is progressing in terms of other radiographic and clinical measures, mSASSS cannot increase past 72. Cross-sectional modeling should attempt to avoid using observations where a score of 72 has been present for many years. A more accurate representation of this subject's disease state is the score preceding complete ankylosis. For example, consider a subject with symptom duration of 30 years who has had an mSASSS of 72 for the last 10 years. The relationship between symptom duration and mSASSS for this subject should be considered as an mSASSS of 72 attained after 20 years, rather than an mSASSS of 72 attained after 30 years. Ignoring the ceiling effect

would underestimate the severity of disease in this subject, making estimates of fixed effects (eg smoking, gender) falsely low.

The mSASSS measure scores twenty four anterior vertebral corners, from the inferior aspect of C2 to the superior aspect of T1, and from the inferior aspect of T12 to the superior aspect of S1. Vertebral corners are often obscured on lateral radiographs leaving an incomplete set of scores. Cervical or lumbar films lacking scores for four or more vertebral corners were deemed invalid. Consistent with mSASSS validation studies, if less than four corners were lacking scores, they were given the mean score for that vertebral segment (cervical or lumbar) (51). A 'contracted mSASSS' (cmSASSS) was used, scoring 0 for normal vertebral corners, erosion, sclerosis or squaring, 1 for a definite syndesmophyte, and 2 for each corner with a total bony bridge (116). Scores for erosion, and squaring were removed as they reduce inter-rater reliability (55, 56). Scores for sclerosis are rare and have poor inter-rater reliability (117). Syndesmophyte formation is not consistently preceded by observed erosion, sclerosis or squaring, making their biological relevance unclear. A recent study found that use of the cmSASSS rather than mSASSS increased complete agreement of scores from 69.7% to 81.4% (116). This study also found that cmSASSS remained highly correlated with mSASSS ($r^2 > 0.96$).

Inter-rater reliability was assessed using a two-way mixed effect Intra-Class Correlation Coefficient (ICC). The ICC is a widely accepted measure of inter-rater reliability suitable for non-normal ordinate data, and is considered the best comparator in radiographic scoring studies (118-120). Longitudinal mSASSS were only available for a few hundred subjects, which would not have provided enough subjects for covariate regression. Scoring of longitudinal mSASSS has been demonstrated to improve inter-rater reliability (55, 59). Sacroiliac joints were scored according to the New York classification (48).

Clinical covariates including co-morbidities had previously been collected by UQDI and TASC. Uveitis, IBD and psoriasis status were physician diagnosed for US subjects and self-reported or physician diagnosed for UK and UQDI subjects. US subjects with uveitis

were ophthalmologist diagnosed, which has a similar rate to self-reported uveitis (121). Data for each covariate were not available for all subjects with mSASSS. Each covariate based analysis used the maximum number of subjects with data on that covariate.

Some covariates vary greatly between clinic visits, and do not lend themselves to cross-sectional analysis. The highly variable covariates ESR, CRP, BASDAI and some symptom scores were included in the univariable analysis as analyses of other cohorts have suggested they have predictive value (65, 92).

2.2.2 Regression Analysis

Regression analysis was used to assess and correct for the influence of measured clinical covariates on the most recent cmSASSS. cmSASSS were not normally distributed due to an extreme left skew.

The NB distribution was the most suitable distribution for modeling cmSASSS. Other distributions commonly used to model count data, such as the Poisson, Gaussian, Gamma, Binomial and exponential distributions are either nested within the NB distribution or do not resemble the distribution of observed data on visible inspection (Fig 2.1). The NB distribution has a high probability when the mean is small, resulting in a left skewed distribution. The most common cmSASSS score was zero, and low cmSASSS were far more common than high cmSASSS, producing a left skewed distribution. The Poisson distribution also includes left skewed distributions, but all Poisson distributions can be produced by the NB function, as they are NB distributions where variance is equal to the mean. Additionally the NB model is favored over Poisson models for analysis of overdispersed data (112, 113). For these reason, models based on the NB distribution was developed. ZINB models have previously been chosen by other researchers for regression analysis of rheumatoid arthritis and mSASSS (79, 109).

Dispersion parameters α and φ were used to assess overdispersion in the NB model. The fit of NB and zero-inflated negative binomial (ZINB) models to the cmSASSS distribution was compared using methods suggested by Hilbe (113) – goodness of fit plots (using Pearson residuals), Vuong's test (a Likelihood Ratio based test used to compare zero-inflated and non-zero-inflated models) and the Akaike Information Criterion (AIC). These goodness-of-fit plots and tests of the NB and ZINB models can be found the Appendix A.4 and A.5. They demonstrate the superior fit of the ZINB model.

Overdispersion in the NB model was assessed by the overdispersion parameter, φ. Unfortunately φ could not be calculated for the ZINB model to allow a straightforward comparison with the value of φ for the NB model. Calculation of φ requires calculation of

the residual deviance of the model, which is then divided by the residual degrees of freedom. Residual deviance cannot be reliably estimated for ZINB as it is a mixture model with a highly non-normal outcome (122). A related and alternative measure of dispersion is the Pearson Dispersion Statistic, PDS (110, 113). This is statistic is calculated by dividing the Pearson Chi Squared Statistic by the residual degrees of freedom of the model, and ideally should be 1.00. In this case, as expected the extra modelling of zeros results in less overdispersion in the ZINB model (PDS = 0.99) compared to the NB model (PDS = 1.03). The superior handling of dispersion by the ZINB model is also reflected by its better results in the goodness-of-fit analysis (Table 2.5, Appendix A.4 and Appendix A.5).

All statistical analyses were performed in R using the packages: *COUNT*, *pscl*, *MASS* and *lmtest* (123-127). *pscl* is the most commonly used ZINB package for R and is considered the most reliable (112, 113). *gamlss* is an R package that is used less commonly for ZINB, but allows random effects (128). Attempts to use *gamlss* with a random effect for Scorer failed due to model non-convergence. A *pscl* ZINB using only fixed effects was used.

Predictors of cmSASSS were assessed using univariable ZINB regression, and were then added by stepwise multivariable regression. The mSASSS scorer of each film was included as a predictor in multivariable analysis, to account for possible minor inter-observer variation. Predictors involving time were allowed to interact with symptom duration to prevent expected collinearity (eg AS symptom duration is believed to interact with smoking duration as they increase equally with age). Covariates which reveal themselves over time were also allowed to interact with symptom duration to account for the cumulative incidence of the covariate over time. For example, AS symptom duration is allowed to interact with uveitis status, a feature of AS with cumulative incidence. An interaction allows uveitis status to have no effect at disease onset (when lifetime uveitis status is inaccurate) and a definite effect after many years of symptoms. Covariates were added stepwise, in order from the most to least significant on univariable analysis.

Superior models were identified by pairwise likelihood ratio tests, AIC and Wald tests on models with the same number of subjects.

A comparison of different measures of smoking was undertaken to identify the best predictor for modeling. Smoking duration, pack year history and 3 level (current/ex-/never) smoking status were not entered into the model at the same time as they are collinear.

Use of an interaction term in R maintains the presence of the fixed effects as well as the interaction term. In this example there would be four effects in the model: symptom duration, uveitis (present/absent), symptom duration with uveitis present and symptom duration with uveitis absent.

One method of limiting the effect of collinearity in a model is to combine highly correlated predictors into a single predictor (110, 112, 129). This approach can involve linear combinations of variables or a regression function that describes their distribution relative to the outcome (110, 130). Horton *et al.* describe using of this method to combine collinear binary, categorical, ordinal and continuous predictors.

As cmSASSS is right censored, use of a model that specifies right censorship may improve fit, including the underestimation of ankylosis. Construction of a right censored ZINB is not currently a feature available in R, and upon perusal of the manuals of SPSS, Stata and SAS this type of model does not appear to be an option usable with the usual ZINB functions. As a future project, a statistician may be able to write their own function that constructs a ZINB that specifies right censorship.

The potential of bias from each scorer was a significant concern during the design of this study. Inter-scorer variation has been a significant concern in previous studies using mSASSS and was a significant predictor in the univariable analysis. The validation studies of mSASSS require calibration of scorers and analysis to assess inter-reader variation. Due to the many scorers across four countries, any inter-reader variation would cause bias in all subjects from an entire site.

The potential influence of scorer was investigated further by generation of confidence intervals for all fixed effects (Table 2.14). The confidence intervals suggested that despite a non-significant effect size in the model, the true effect size of certain scorers may be significant. Bias attributed to scorer may also result from site specific factors. Scorers 6 and 7 scored films from the same hospital, but all other scorers work at different hospitals across the United States, Canada and the United Kingdom. Local factors including genetic variation, radiographic methods, occupational or other environmental factors may vary between these sites.

2.2.3 Sensitivity Analysis of the Cross-Sectional Model

The main goals of sensitivity analysis are to identify if variability and uncertainty of input parameters have significant and relevant effects, (131) and to quantify the requirements of an unmeasured confounder large enough to alter the conclusions of a study (132).

Terminology for sensitivity analyses varies between authors, due to the overlapping nature of statistical techniques each uses. This study used a combination of techniques which are described by various authors as one-at-a-time analysis, local sensitivity analysis, range sensitivity analysis, confounder analysis, regression sensitivity analysis and graphical analysis.

Local sensitivity analysis can be performed by varying one predictor at a time by a plausible range of variation or within the known range of input values (133). Range sensitivity analysis examines model sensitivity across the entire range of input variable values.

Local sensitivity analysis was performed by simulation, using either all observed data or a reference dataset of all possible combinations of observed values. For example, considering only 3 level smoking status, scorer and gender, there were 3 smoking statuses x 7 scorers x 2 genders = 42 possible combinations → 42 entries in the reference

population, which cover the entire probability space of observed data. The ZINB model was run on this reference population, generating the entire range of possible predicted cmSASSS.

To generate confidence intervals for model predictors, observed data (not reference data) underwent bootstrap resampling to create a simulated population size of 10 000, and the model rerun. Bootstrapped confidence intervals are one method of demonstrating that a predictor has a significant effect on the model, and of quantifying the uncertainty of its effect size.

The relative contributions of predictors to model outcome were assessed with r^2 and pseudo-r^2 values, calculated by linear regression of the model predicted value against each predictor.

The effect of possible unknown confounders was performed by simulating the effect of adding a synthesised confounding covariate. Two synthesised confounders were made: a present/absent factor and a normally distributed continuous covariate. Both were created from normalised ranked values of each subject's New York sacroiliac joint score, a metric highly correlated with mSASSS. The confounders were given a range of effect sizes (0-1 standard deviation per unit), and the present/absent factor a range of prevalence (0-100%), and the effect on predicted cmSASSS was simulated.

2.3 Results

2.3.1 TASC Cohort Characteristics

Clinical data on gender and symptom duration were available on 1392 subjects with valid cmSASSS (Table 2.1).

Table 2.1. Available covariate data

Covariate	Subjects with covariate and a valid mSASSS	Subjects with Gender, symptom duration, valid mSASSS and covariate
Film scorer	1464	1392
Gender	1417	1392
Symptom duration	1411	1392
BASDAI ever	1006	986
BASFI ever	1176	1143
Age of symptom onset	1231	1204
3 level Smoking (ex-, current, never)	1135	1117
Age	1228	1207
Smoking duration	1135	1113
Pack year history	908	896
ESR ever	384	371
CRP ever	361	354
IBD	166	165
Psoriasis	538	536
Uveitis	538	536
Mean recorded Visual-Analogue Scale (VAS) stiffness*	568	548
Patient Stiffness*	580	557
Exercise Number/wk*	567	547
Exercise minutes/wk*	557	538

Regular Back Exercises*	566	546
Back Exercises days per wk*	566	546

* See appendix A.1 for description

The cohort had similar rates of male gender, uveitis, IBD and psoriasis to other AS cohorts (Table 2.2). The average subject was aged in their forties, had symptoms for twenty-one years, and was an ex-smoker. There is a trend towards a lower age of symptom onset lower in males than females, with overlapping 95% confidence intervals (p=0.02 by t test, male mean 24.7, 95% CI 24.28-25.13, female mean 25.6, 95% CI 24.95-26.30).

Table 2.2. Cohort Profile

Covariate	TASC Cohort	OASIS Cohort (at baseline) (64)	Other cohort meta-analysis (100)
Subjects	1392	217	
Gender (% male)	74.4%	70%	
Mean symptom duration (yrs)	21.5	11	
Mean age of symptom onset	25.0		
Mean follow up (yrs)	2.7	7.9	
Mean age	47.5	43	
Mean latest mSASSS	10.9	11.6	
Mean sacroiliac score	3.2		
Pack year history	Mean 6.7* Median 0		
3 Level Smoking	16% Never 29% Current 54.8% Ex-		
Uveitis ever	32.6%		25.8%
Uveitis in subjects with symptom duration 10-15	26.0%		23.3%

yrs			
Uveitis in subjects with symptom duration > 20 yrs	35.6%		38.5%
Diagnosed IBD	8.1%		6.8%
Diagnosed psoriasis	11.3%		9.3%

* use of mean is inappropriate given the non-normal distribution (Figure 2.3B)

The contracted mSASSS distribution was left skewed with 509 scores of zero and cmSASSS range 0 to 48.

Figure 2.1. Distribution of cmSASSS

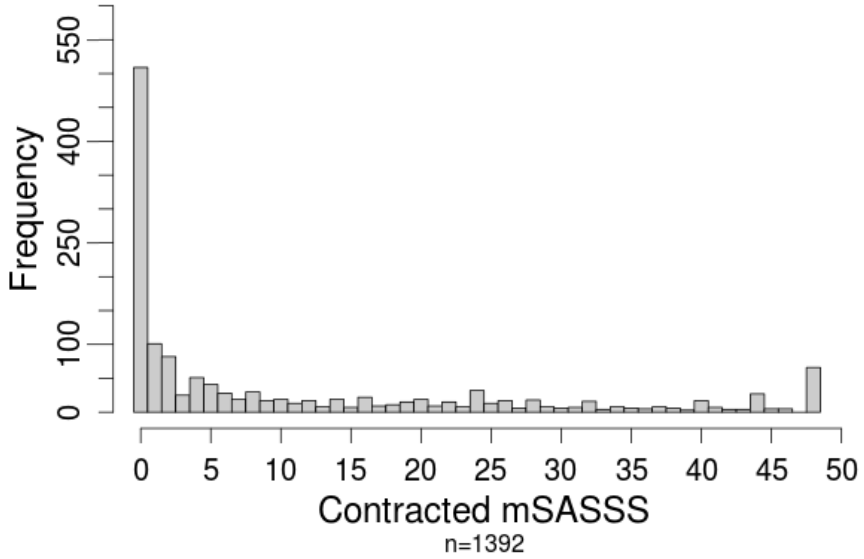


Figure 2.2. cmSASSS vs symptom duration. Males in blue, females in red. cmSASSS increases with symptom duration, but large variation exists. Use of correlation (Spearman's rho = 0.49) is poorly descriptive – it belies the abundance of scores of zero.

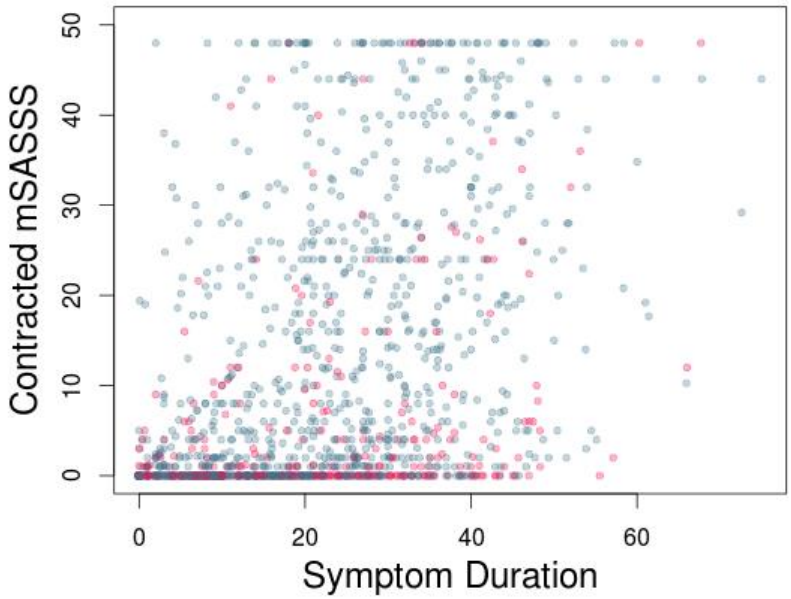


Figure 2.3. Histograms of symptom duration and smoking duration. There is skew towards lower symptom duration and lower smoking duration. Notably only 16% of subjects self-report never smoking, yet 63% report a smoking duration of 0 years. This suggests most ex-smokers have a very limited smoking duration.

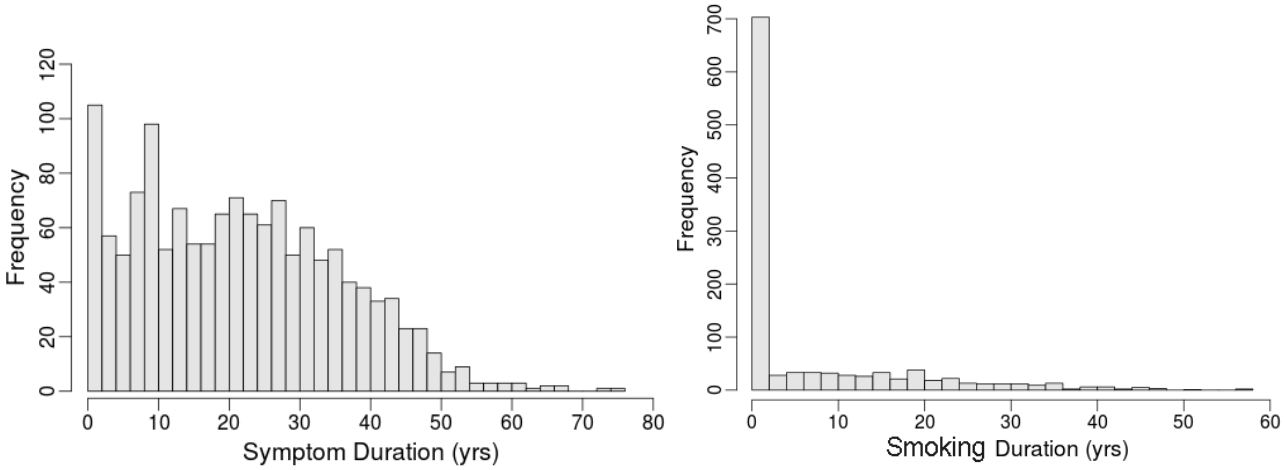


Figure 2.4. Boxplot comparing cmSASSS by gender. Across symptom duration, males have higher cmSASSS.

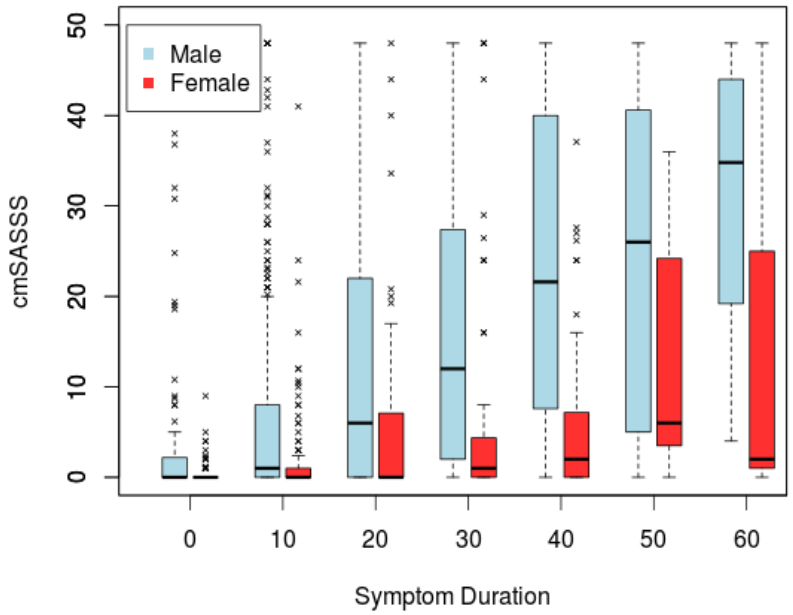
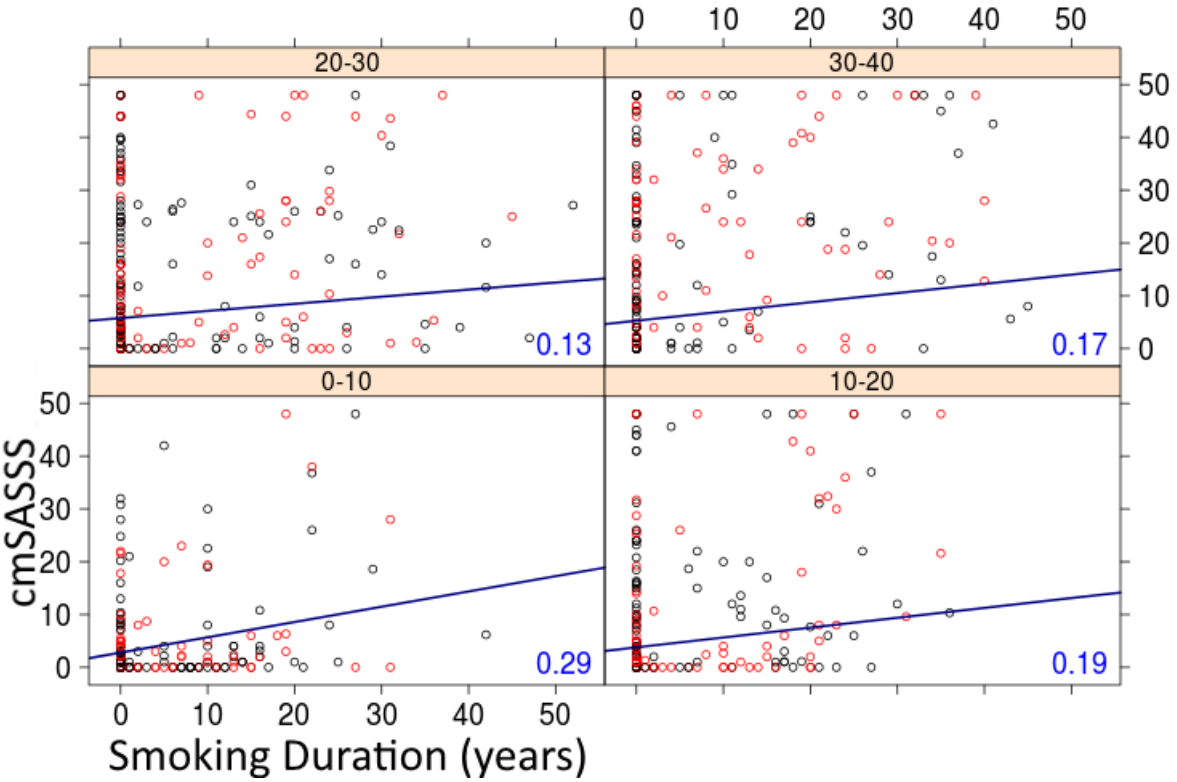


Figure 2.5. cmSASSS vs Smoking Duration across four bins of symptom duration. Each range of smoking duration has a different panel. Males are represented by black circles, females by red. For each panel, a blue line represents a simple linear model fit to cmSASSS vs smoking duration (gradient at the bottom right of each panel). cmSASSS appears to be proportionate to smoking duration across symptom duration.



Many observations of zero exist for smoking duration which was the chosen value for subjects who have never smoked. These 'never smoker' subjects were included in the analysis to allow for a spectrum of smoking exposure from never smoked to heavy exposure. The use of zero for never smokers is a commonly used value for smoking history (eg pack year history, cigarette year history) (134).

The use of zero for never smokers assumes a quantitative rather than qualitative relationship, where the effect size per year of smoking is equal (ie a linear relationship). The effect size for smoking duration is averaged so that the effect of smoking for 0 vs 10

years is the same as smoking for 40 vs 50 years. If the effect size of smoking is not constant, use of a smoking duration of zero for never smokers may under or overestimate the effect of smoking for some individuals. Importantly in this case, plots of cmSASSS vs smoking duration do not demonstrate any definite non-linear relationship with the cmSASSS vs smoking gradient remaining similar across symptom durations >10 years (Fig 2.5). The gradient of the 0-10 smoking duration bin is higher, which would suggest that the effect size between never smokers and low duration smokers is higher than the gradient for short vs long duration smokers.

Cervical cmSASSS scores contribute more to total cmSASSS than lumbar scores (Table 2.3).

Table 2.3. Correlation between observed values

Observation 1	Observation 2	Spearman Correlation
Cervical cmSASSS	Lumbar cmSASSS	0.61
Cervical cmSASSS	Total cmSASSS	0.91
Lumbar cmSASSS	Total cmSASSS	0.88
Mean BASFI	Total cmSASSS	0.23
Total mSASSS	Total cmSASSS	0.99
Cervical mSASSS	Total cmSASSS	0.98
Lumbar mSASSS	Total cmSASSS	0.99
VAS Stiffness	Total cmSASSS	0.27
Early Morning Stiffness	Total cmSASSS	0.14
NY Sacroiliac Score	Total cmSASSS	0.54

2.3.2 Inter-rater Reliability

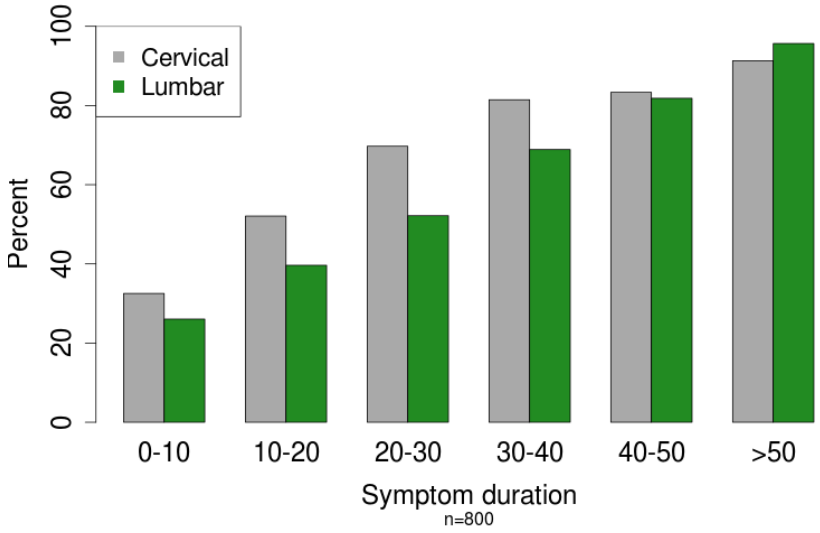
The ST-UQDI ICC was 0.968 (95% CI 0.935 – 0.984, $p=2.06 \times 10^{-20}$) and the ST-TASC ICC was 0.985 (95% CI 0.965-0.994, $p=3.56 \times 10^{-17}$). These ICC indicate good to excellent reliability (135). The inter-rater reliability of scorers MAB, TJL, MMW and MHW were found to be high in a previous study that used the same TASC reference films (116).

2.3.3 Cervical and Lumbar score comparison

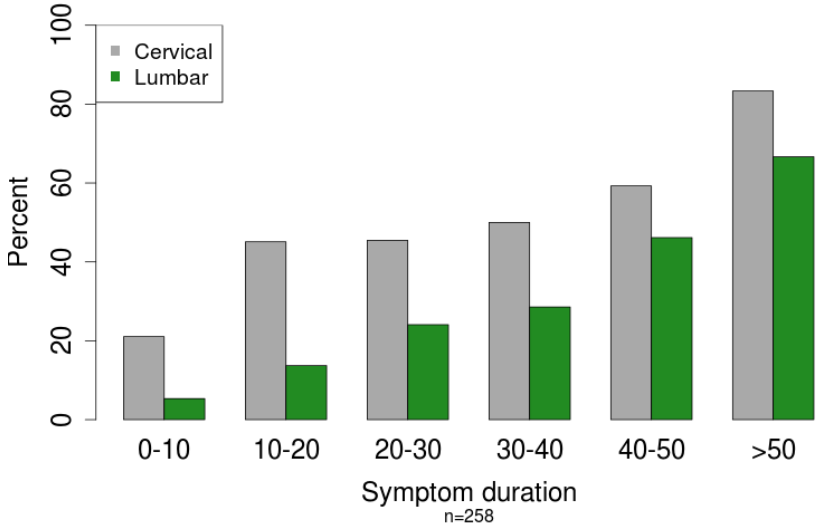
In both genders, across almost all points of symptom duration, syndesmophyte presence was more common in the cervical spine (Figure 2.6). Presence of lumbar syndesmophytes was far more common in males.

Figure 2.6. Frequency of cervical and lumbar syndesmophyte across symptom duration.

A) Males



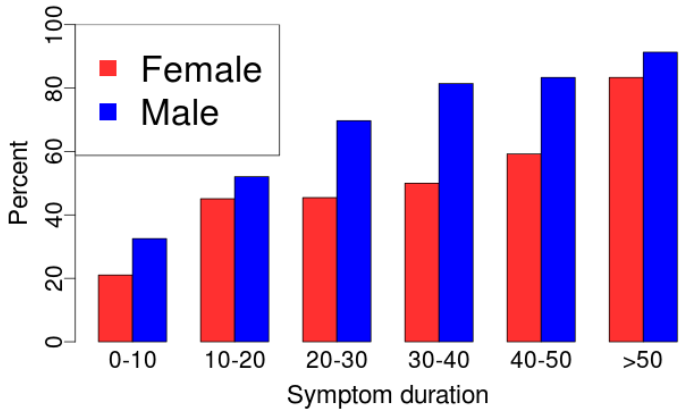
B) Females



Males are more likely to have syndesmophytes in the cervical or lumbar spine. Males have a markedly higher rate of lumbar syndesmophyte presence (Figure 2.7).

Figure 2.7. Presence of cervical and lumbar spine syndesmophytes by gender.

A) Cervical Spine



B) Lumbar Spine

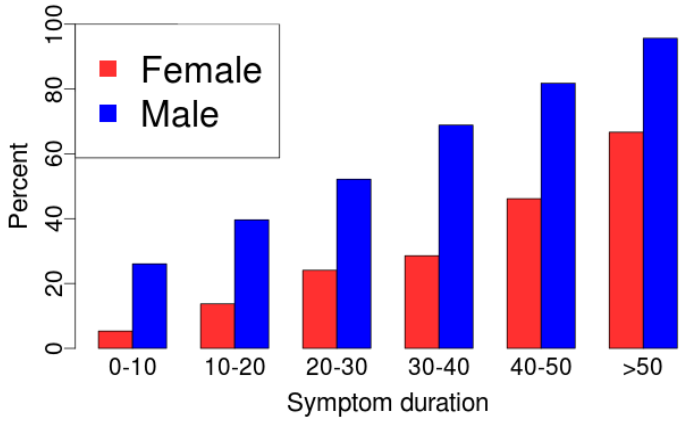


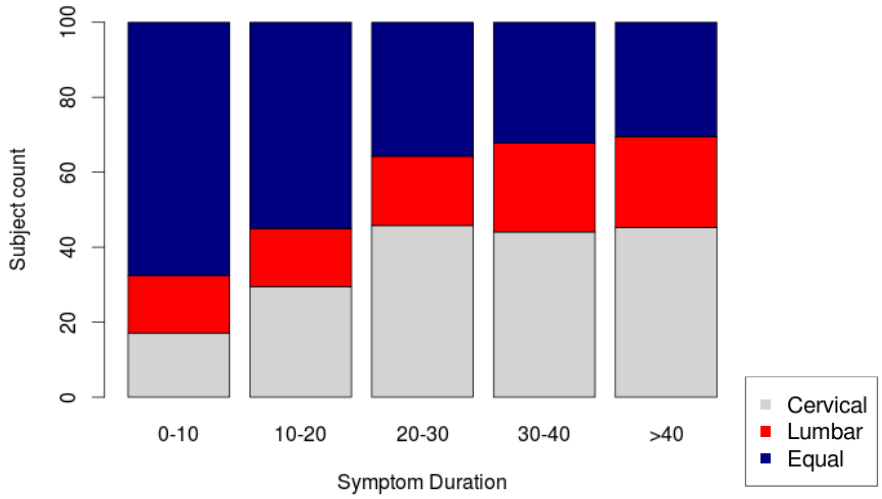
Table 2.4. *Inter-gender comparison*. Comparison of results from Figures 2.7A and 2.7B by Chi Squared testing. Males have more severe cervical and lumbar disease at almost all points of symptom duration.

Symptom duration	**Cervical syndesmophyte presence is higher in males (p value)**	**Lumbar syndesmophyte presence is higher in males (p value)**
0-10	0.03	2.1×10^{-5}
10-20	0.3	2.3×10^{-5}
20-30	1.4×10^{-4}	4.1×10^{-4}
30-40	3.2×10^{-4}	9.3×10^{-4}
40-50	0.04	7.25×10^{-5}
>50	Not calculated as female n=4, male n=21	

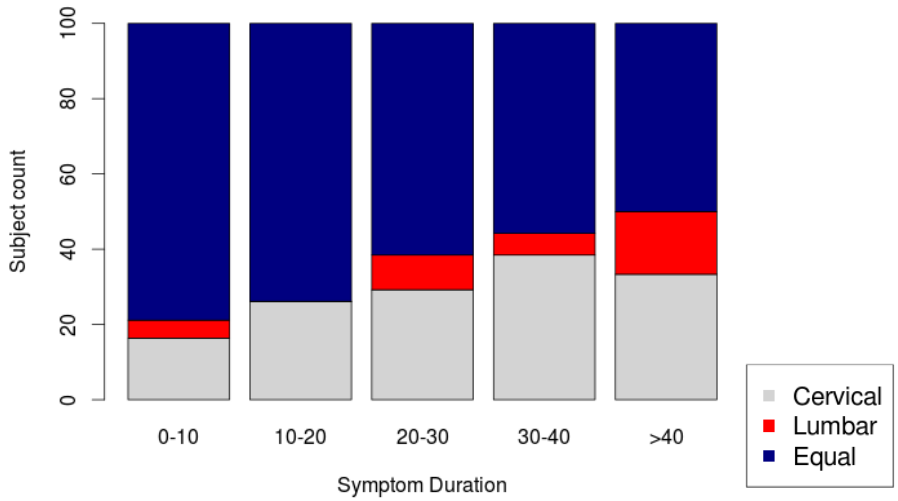
Cervical predominant disease, as defined by previous authors was more common in males than lumbar or equal predominance after forty years of symptoms (61). Most females did not display a cervical or lumbar predominance.

Figure 2.8. Predominance of cervical and lumbar disease by gender. Predominant disease is defined by a cervico-lumbar difference >2 cmSASSS. The proportions of cervical and lumbar predominant disease are different between genders at all points of symptom duration ($p<0.02$ in five separate Chi Squared tests, for five symptom duration bins).

A) Males



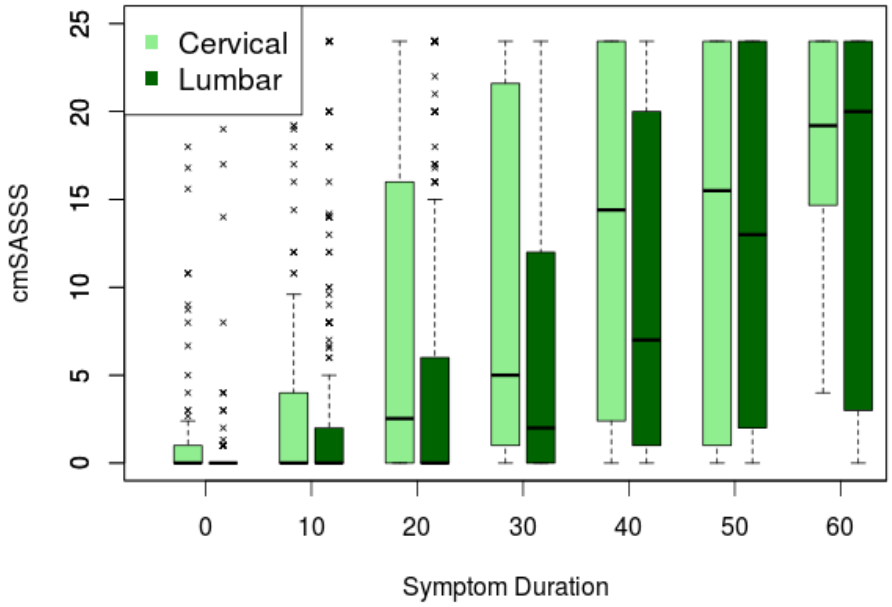
B) Females



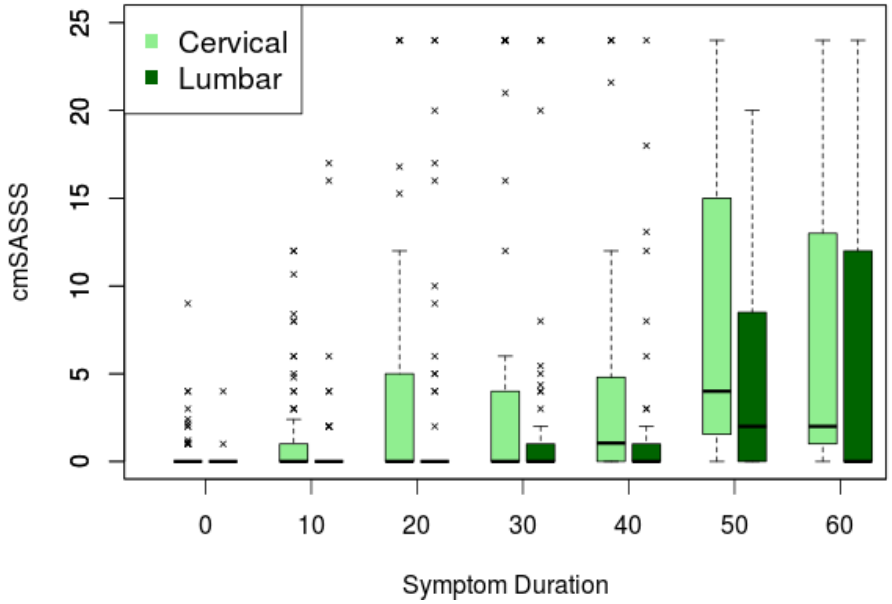
Over a lifetime, cervical cmSASSS generally rise before lumbar cmSASSS. The rise of cmSASSS in females is mild compared to males (Figure 2.9).

Figure 2.9. Distribution of cervical and lumbar mSASSS by gender. Early in the disease course, cervical cmSASSS are markedly higher than lumbar cmSASSS, in both genders.

A) Males

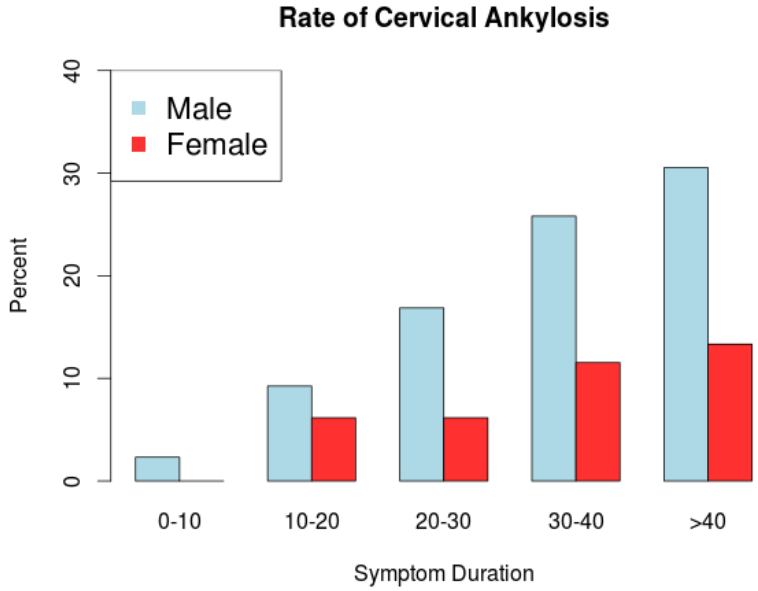


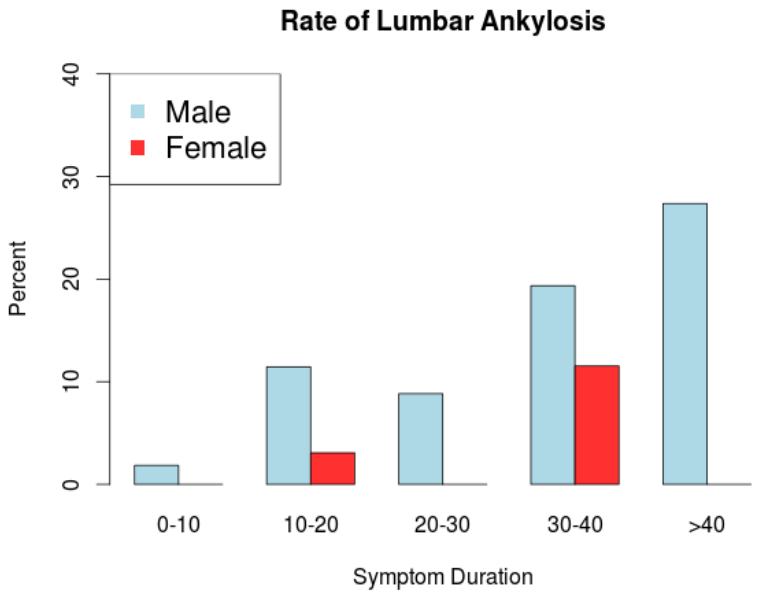
B) Females

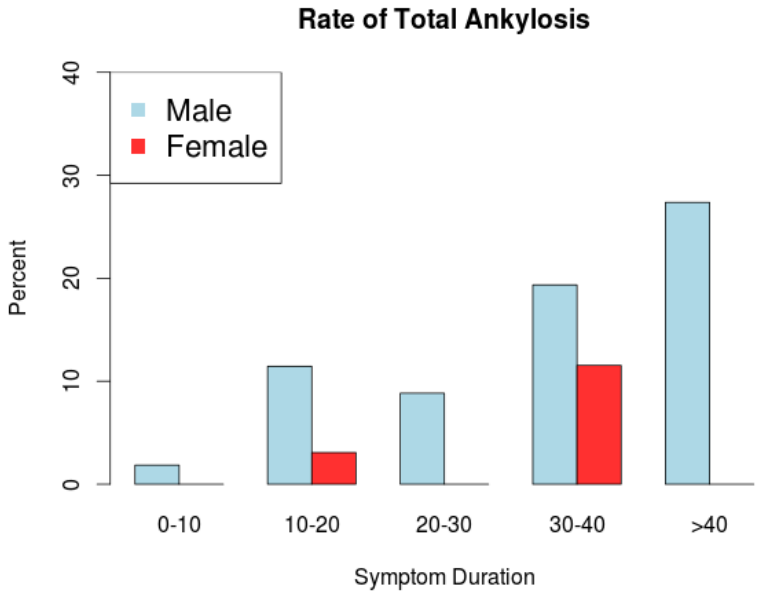


After forty years of symptoms, complete lumbar ankylosis is rare in females, but occurs in 30% of males. Complete cervical ankylosis also occurs in ~30% of males after forty years, and 13% of females (Figure 2.10).

Figure 2.10. Rates of ankylosis across symptom duration. Subject numbers: 973 males and 316 females.



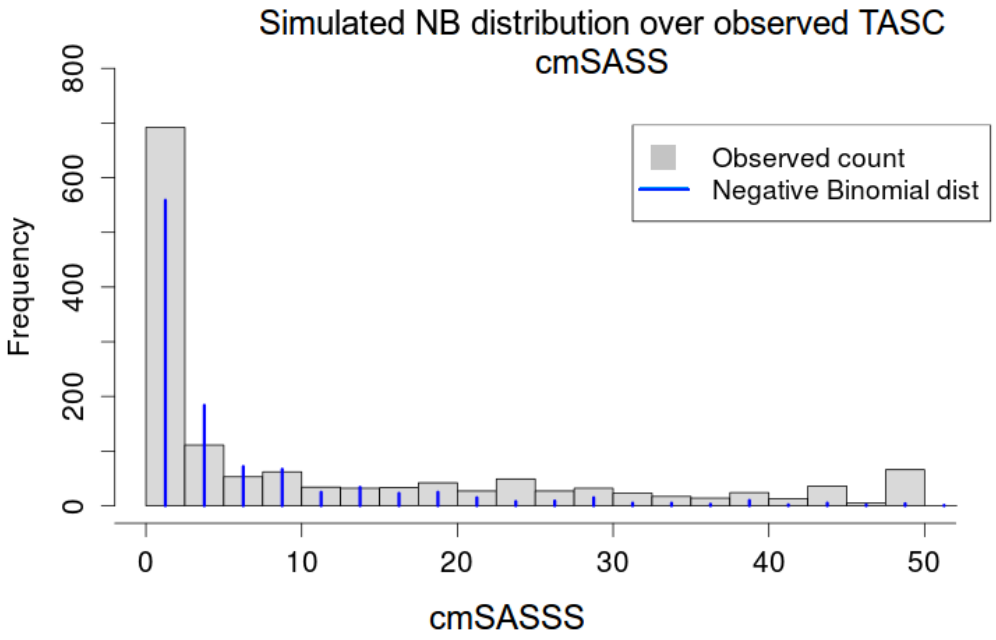

Rate of Lumbar Ankylosis
Male
Female
Percent
0
10
20
30
40
0-10
10-20
20-30
30-40
>40
Symptom Duration


Rate of Total Ankylosis
Male
Female
Percent
0
10
20
30
40
0-10
10-20
20-30
30-40
>40
Symptom Duration

2.3.4 Regression Analysis

The cmSASSS distribution resembled the negative binomial distribution (Figure 2.11).

Figure 2.11. A negative binomial distribution overlaid on observed cmSASSS counts. Notably cmSASSS approaching 48 (complete ankylosis) were underestimated by this distribution. The there is an extreme observed cmSASSS count of zero – zero inflation. The NB distribution under-predicts the observed number of zeros.



A basic negative binomial (NB) model using symptom duration and gender as predictors of cmSASSS was fitted to the data. It displayed significant overdispersion, demonstrated by dispersion parameters: $\alpha = 2.64$ and $\varphi = 1.067$ (calculations in Appendix A.2) (112, 113). Correctible causes of real and apparent overdispersion, such as missed predictors/covariates, outliers, missed predictor interactions, or a predictor or response requiring transformation were investigated with multiple NB models. All displayed overdispersion. Overdispersion can also be caused by excessive counts of zero, which are present in this case. Overdispersion due to excessive counts of zero can be addressed with a zero-inflated negative binomial model, which is a two component mixture model of the negative binomial distribution and a zero model that estimates how many

counts of zero are likely due to covariate influence (112, 113). In this case, many cmSASSS of zero are attributable to low symptom duration, making this model biologically relevant. Comparison testing, including Vuong's test, AIC and goodness-of-fit plots demonstrate the superior fit of the ZINB model over the NB (Table 2.5. See Appendix A.4 and A.5 for goodness-of-fit plots).

The cause of the real overdispersion was most likely the excessive counts of zero. This particular cohort and this condition include many patients with mSASSS of zero. The only other relevant cause of real overdispersion, data clustering from population heterogeneity, is not present in this case. There is no evidence of clusters of scores representing sub-populations in the cmSASSS histogram (Fig 2.11).

ZINB and NB models are based on the distribution seen in Figure 2.11, which decreases at higher values. Thus both model types are likely to underestimate the number of individuals with complete ankylosis. This is seen in Figure 2.12, where predicted values are significantly lower than observed values. For the purpose of this modelling, this underestimation is acceptable, so long as the and the predicted scores represent the relative differences between the severity of individuals.

Figure 2.12. The ZINB distribution overlaid onto observed cmSASSS. The ZINB distribution allows for a higher count of zero.

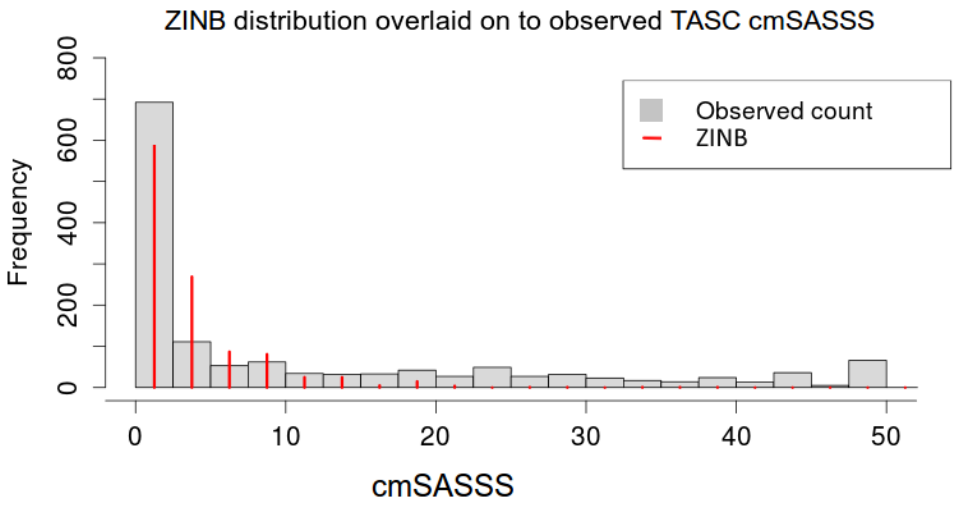
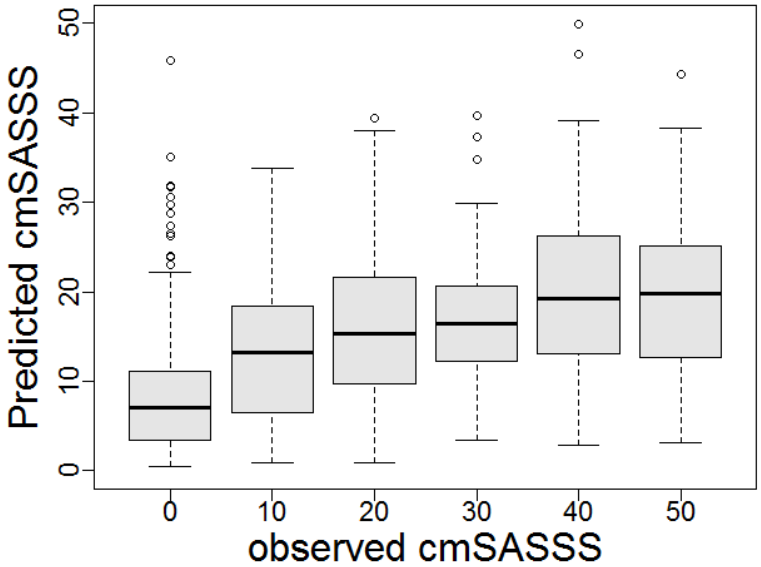
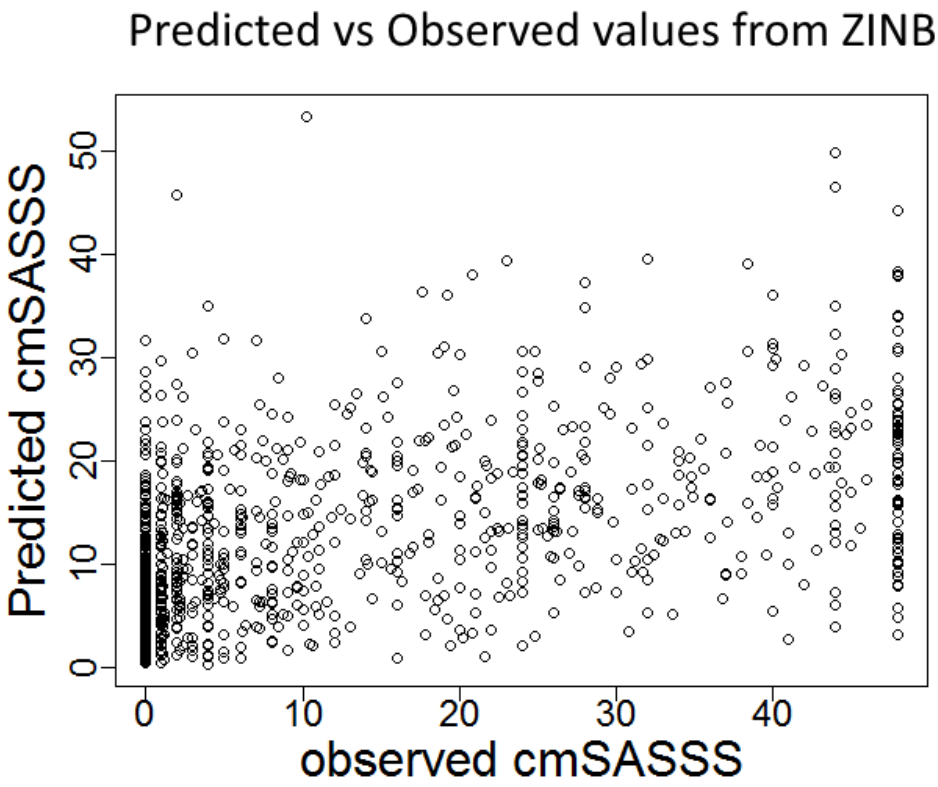


Figure 2.13. Predicted vs observed values from the ZINB model – Boxplot and scatterplot. Predicted values increase as observed values increase, but their size is less than expected. For example, the median predicted cmSASSS for an observed cmSASSS of 20 is 15.





ZINB and NB models are based on the distribution seen in Figure 2.11, which decreases at higher values. cmSASSS is interval censored, so individuals with severe disease become ankylosed and never change their score, causing a 'ceiling effect'. This results in a peak in the histogram at the maximum cmSASSS of 48. To minimize the ceiling effect, if available, the observation immediately before complete ankylosis was used instead of the latest score of 48. Unfortunately no statistical distribution that I have read about has a peak at extreme low and high values. The ideal distribution to model this data might be a distribution which has a peak at zero and at 48. I have not encountered such a distribution. In my reading of modeling count data, apart from longitudinal modeling, I have seen no more appropriate distribution or method than ZINB model, and this method has been used by previous authors for mSASSS (79) and other musculoskeletal radiology papers (109). Consequently we must accept that this model will underestimate the number of individuals with complete ankylosis, and keep this in mind when interpreting the result.

Table 2.5. *Vuong's test and AIC favour the ZINB model.* Lower AIC indicates superior fit.

Model	Vuong's test	AIC
NB	ZINB superior $p < 2.2 \times 10^{-16}$	8358.96
ZINB		8138.72

Univariable regression analysis was used to assess the predictive value of recorded covariates (Tables 2.6, 2.7). *pscl* reports ZINB results for both the zero and count components of the model. The 'count model' assesses which covariates predict higher cmSASSS other than zero (ie 1 vs 48). The 'zero model' assesses which covariates predict scores of zero. Thus a covariate which predicts higher cmSASSS should have a positive effect size for the count model, and a negative effect size for the zero model. Covariates which are only significant in the zero model may only be significant because of this two component approach, or may truly predict scores of zero without influencing further syndesmophyte formation, which is of little clinical relevance.

Univariable analysis indicates scorers 4 and 5 scored higher mSASSS (Table 2.6) than other scorers. Symptom duration, male gender and pack year history predicted higher cmSASSS. Current smoking and age predicted higher cmSASSS, and were collinear with smoking duration and symptom duration respectively. The effect of gender on cmSASSS was similar at all levels of symptom duration, as indicated by the absence of significant interaction between gender and symptom duration. Cross-sectional measures of disease activity, including ESR, CRP and mean recorded BASDAI were not significant, while BASFI had a significant association. Two stiffness related scores had significant p-values, though one of these had confidence intervals overlapping zero. Higher age of symptom onset was only significant in the zero model.

Table 2.6. ZINB univariable analysis of film scorer influence. All values are relative to Scorer 2 who scored the most films used in this study, 607.

Scorer	Count Model			Zero Model			Films
	Effect Size	95% CI	P value	Effect Size	95% CI	P value	
1	-0.024	-0.33, 0.29	0.88	-0.52	-1.08, 0.04	0.07	107
2	-	-	-	-	-	-	607
3	-0.021	-0.23, 0.19	0.84	-0.66	-1.04, -0.28	7.5×10^{-4}	302
4	0.34	0.10, 0.58	5.3×10^{-3}	-1.08	-1.62, -0.55	7.6×10^{-5}	173
5	0.48	0.13, 0.83	6.5×10^{-3}	-1.71	-2.87, -0.55	3.87×10^{-3}	63
6	0.54	-0.14, 1.22	0.12	-1.41	-3.35, 0.53	0.15	15
7 (ST)	-0.15	-0.43, 0.12	0.29	-0.89	-1.51, -0.28	4.5×10^{-3}	125

Table 2.7. Results of covariate univariable regression analysis. Ordered by count model p value. Signficant p-values are shaded.

Covariate	Count Model			Zero Model		
	Effect Size	**95% CI**	**P value**	**Effect Size**	**95% CI**	**P value**
Symptom duration	0.03	0.02, 0.03	$<2x10^{-16}$	-0.09	-0.10 -0.07	$<2x10^{-16}$
Age	0.03	0.02, 1.54	$<2x10^{-16}$	-0.09	-0.1, -0.07	$<2x10^{-16}$
Male Gender	0.63	0.42, 0.84	2.98×10^{-9}	-1.28	-1.59, -0.97	1.06×10^{-15}
Smoking duration (yrs)	$1.6x10^{-2}$	0.01, 0.02	5.29×10^{-6}	-0.08	-0.1, -0.05	3.76×10^{-8}
Mean recorded VAS stiffness	0.01	0.01, 0.02	2.35×10^{-5}	-0.01	-0.02, -0.01	3.6×10^{-4}
Mean recorded BASFI	6.1×10^{-3}	$3x10^{-3}$, 0.01	8.35×10^{-5}	-0.01	-0.02, -0.01	$1.37x10^{-4}$
Pack year history (20 / pack)	0.01	$4.9x10^{-3}$, 0.02	5.2×10^{-4}	-0.14	-0.23, -0.06	$7.4x10^{-4}$
Current smoker vs never	0.28	2.1×10^{-3}, 0.55	0.048	-0.62	-1.12, -0.12	0.015
Mean recorded BASDAI	-0.03	-0.06, $2.2x10^{-3}$	0.07	0.05	-0.01, 0.11	0.089
Age of symptom onset	6.85×10^{-3}	2.1×10^{-3}, 0.02	0.14	-0.02	-0.04, -3.9×10^{-3}	0.02
Early Morning Stiffness	5×10^{-4}	0, 0	0.27	-2.5×10^{-3}	0, 0	5.9×10^{-3}
Regular Back Exercises	-0.21	-0.6, 0.17	0.28	-0.12	-0.65, 0.4	0.64
Exercise Number/wk	0.02	-0.03, 0.07	0.35	-0.06	-0.14, 0.02	0.17
Ex-smoker vs never	0.12	-0.15, 0.38	0.39	-0.14	-0.57, 0.3	0.53
Latest ESR	4×10^{-3}	-0.01, 0.01	0.4	-0.02	-0.05, 0.01	0.2
IBD	0.19	-0.52, 0.9	0.6	-1.4	-4.25, 1.42	0.32
Back Exercises: days per wk	0.02	-0.07, 0.1	0.67	-0.03	-0.15, 0.08	0.57

Latest CRP	2×10^{-3}	2.24, 2.75	0.68	-0.11	-0.29, 0.06	0.21
Uveitis	-0.03	-0.22, 0.16	0.74	0.07	-0.27, 0.4	0.69
Psoriasis	-0.03	-0.24, 0.19	0.81	0.21	-0.17, 0.59	0.29
Exercise minutes/wk	-1.8×10^{-4}	0, 0	0.88	-2×10^{-3}	-0.01, 0	0.35

Covariates with univariable count model p-value < 0.1 were progressed to multivariable analysis. Age was not included as it is collinear with and of less clinical relevance than symptom duration. VAS stiffness and BASFI were not progressed to multivariable analysis as they are outcomes or results of higher cmSASSS and not biological processes. Including them in the model would decrease variance attributed to other predictors. Age of symptom onset was only significant in the zero model, but has been found to predict higher mSASSS in previous studies so was progressed to multivariable analysis (71).

Some univariable results of borderline significance did not produce results consistent with known pathophysiology. BASDAI had univariable p-values of borderline significance, from 0.05 to 0.1, with a negative effect size in the count model, and an effect size in the zero model with a confidence interval overlapping zero. The negative effect size suggests higher BASDAI predicts lower cmSASSS, which is not consistent with previous studies that suggest higher BASDAI, ESR and CRP predict radiographic severity (92, 93). This effect might be explained by the protective effect of NSAIDs and TNF inhibitors, which are more likely to be initiated in subjects with high BASDAI.

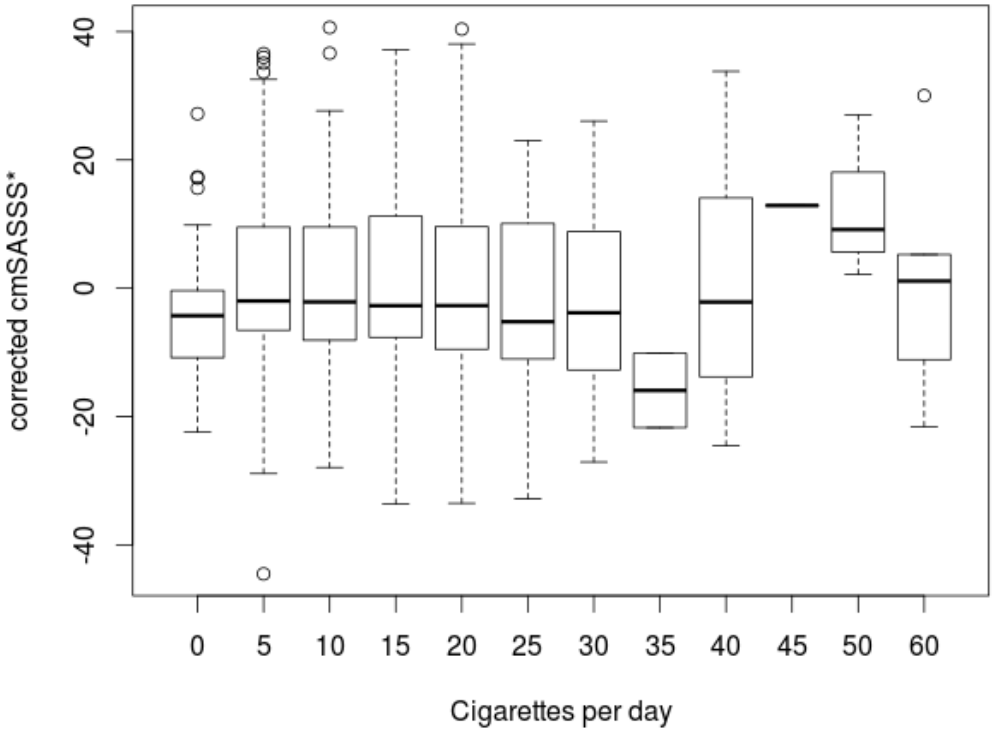
Comparison between no smoking status, 3 level smoking status, pack year history and smoking duration using model AICs indicate that smoking duration is the most predictive smoking metric (Table 2.8). Pack year history is calculated by: smoking duration multiplied by reported cigarette intake per day ÷ 20. To analyse whether cigarette intake per day holds any predictive value, it was assessed in the multivariable model in place of pack year history, and was found to be non-significant (p=0.55). In terms of overall model fit, there is little difference between the smoking metrics listed in Table 2.8 (<1% of AIC). The interaction between smoking duration and symptom duration was also significant with a

negative effect size, suggesting that as symptom duration and smoking duration increase, the effect size of smoking duration decreases very slightly ($p = 0.045$, effect size -3.6×10^{-4}, 95% confidence intervals -7.1×10^{-4}, -7.6×10^{-6}). Only the most predictive smoking metric, smoking duration, was progressed to multivariable analysis.

Table 2.8. *Multivariable analysis of smoking measures*. A multivariable model corrected for symptom duration, gender and scorer. Each smoking measure was analysed in a separate model, and each model had the same 890 subjects, each with valid entries for smoking duration, 3 level smoking and pack year history.

Smoking metric	**Count Model**			**Zero Model**			**AIC**
	Effect size	**95 % CI on effect size**	**P-value**	**Effect size**	**95 % CI on effect size**	**P-value**	
No smoking status	-	-	-	-	-	-	5459
3 level smoking: ex	0.24	-0.27, 0.75	0.35	0.14	-0.77, 1.06	0.75	5449
3 level smoking: current	0.33	-0.2, 0.86	0.22	-1.31	-2.32, -0.32	9.9×10^{-3}	
Pack year history	0.01	-0.01, 0.04	0.19	-0.13	-0.23, -0.04	7.1×10^{-3}	5433
Smoking duration	0.024	0.01, 0.04	9.4×10^{-4}	-0.08	-0.12, -0.04	1.35×10^{-5}	5422

Figure 2.14. Analysis of pack year history. After correcting for smoking duration, gender and symptom duration, cigarettes per day are not associated with cmSASSS.



Gender, symptom duration, smoking duration and age of onset remained significant on stepwise multivariable analysis. Mean recorded BASDAI was not. Symptom duration was calculated by subtracting current age from age of symptom onset, so age of symptom onset was allowed to interact with symptom duration. The modelling software would not allow symptom duration to interact with smoking duration and age of onset at the same time, so two models, each interacting with one only were modelled. Both found age of onset to be a significant predictor of cmSASSS, and the age of onset – symptom duration interaction model found no significant interaction between the two. Surprisingly, age of onset was not significant in the count model upon univariable analysis, but became significant upon multivariable analysis. This is likely because of its relationship to symptom duration. Explanations for its newfound significance upon multivariable analysis are that it may be a genuine effect of small size, only detectable once symptom duration is corrected for, or it may be an artefact of this correction.

Table 2.9. *Stepwise selection.* Multivariable model comparison

Model: cmSASSS vs	New Covariate	Count Model			Zero Model		
		Effect size	95 % CI on effect size	P value	Effect size	95 % CI on effect size	P value
Symptom duration + Scorer	Symptom Duration	0.03	0.02, 0.03	4.73×10^{-14}	- 0.10	-0.12, -0.08	$<2.2 \times 10^{-16}$
Symptom duration + Scorer + Gender	Gender	0.65	0.44, 0.86	2.24×10^{-9}	-1.05	-1.44, -0.67	9.97×10^{-8}
Symptom duration * smoking duration + Scorer + Gender	Smoking duration	0.02	0.01, 0.04	1.53×10^{-3}	-0.1	-0.14, -0.06	4×10^{-7}
Symptom duration * smoking duration + Scorer + Gender + BASDAI	Mean BASDAI	0.02	-0.01, 0.05	0.25	-0.07	-0.15, 2.6×10^{-3}	0.06
Symptom duration * smoking duration + age of onset + Scorer + Gender	Age of onset	0.02	3.8×10^{-3}, 0.04	0.017	-0.05	-0.08, -0.02	1.7×10^{-3}
Symptom duration * age of onset + smoking duration + Scorer + Gender	Age of onset	0.02	-0.1, -0.03	0.03	-0.06	2.2×10^{-3}, 0.04	3×10^{-4}

Table 2.10. Age of onset model comparison

Model: cmSASSS vs	New Covariate	Count Model			Zero Model		
		Effect size	95 % CI on effect size	P value	Effect size	95 % CI on effect size	P value
Symptom duration * smoking duration + age of onset + Scorer + Gender	Age of onset	0.02	3.8×10^{-3}, 0.04	0.017	-0.05	-0.08, -0.02	1.7×10^{-3}
Age * smoking duration + age of onset + Scorer + Gender	Age of onset	-5.1×10^{-3}	-0.01, 3.5×10^{-3}	0.24	0.04	0.02, 0.06	1.2×10^{-3}

The positive effect size of age of symptom onset in the count and zero models suggests that a later age of onset predicts a higher cmSASSS. Proof of this relationship requires further investigation given that symptom duration and age of onset are known to interact. To test for a symptom duration – age of onset interaction, a multivariable model using age instead of symptom duration was run. When age is used instead of symptom duration, age of onset is no longer significant. A univariable model, comparing five age of onset 'bins' (<16, 16-20, 20-30, 30-40, 40-50 years) to cmSASSS found no bin to be significant ($p>0.05$), and a multivariable model, correcting for symptom duration did not either. However, if the dataset is limited to subjects with symptom duration longer than twenty years, onset before age 16 becomes significant in univariable and multivariable models ($p=0.002$ and $p=0.0002$ respectively), suggesting there is a small yet significant effect of early age of symptom onset, which has a long lead time, and which is detectable without correction for symptom duration.

Alternate explanations for the positive effect size of age of onset is that older age of onset subjects may have more syndesmophytes at baseline, or may more have recall bias. Initial cmSASSS is higher in subjects with a higher age of onset (Figure 2.14). This is evident upon linear regression analysis and t testing between bins. Linear regression demonstrates the higher initial cmSASSS in older subjects with symptoms for less than 5 years : effect size 0.14 (95% CI 0.06, 0.22, p value 5×10^{-5}). Comparison of decade of onset bins (Table 2.11) demonstrates higher cmSASSS at later ages of onset in recent onset AS.

Figure 2.15. Plots of first recorded cmSASSS vs age of onset. Limited to subjects with symptom duration < 5 years. Notably, some subjects are recorded as having age of onset > 50 years which is highly unusual in AS and suggests earlier asymptomatic disease.

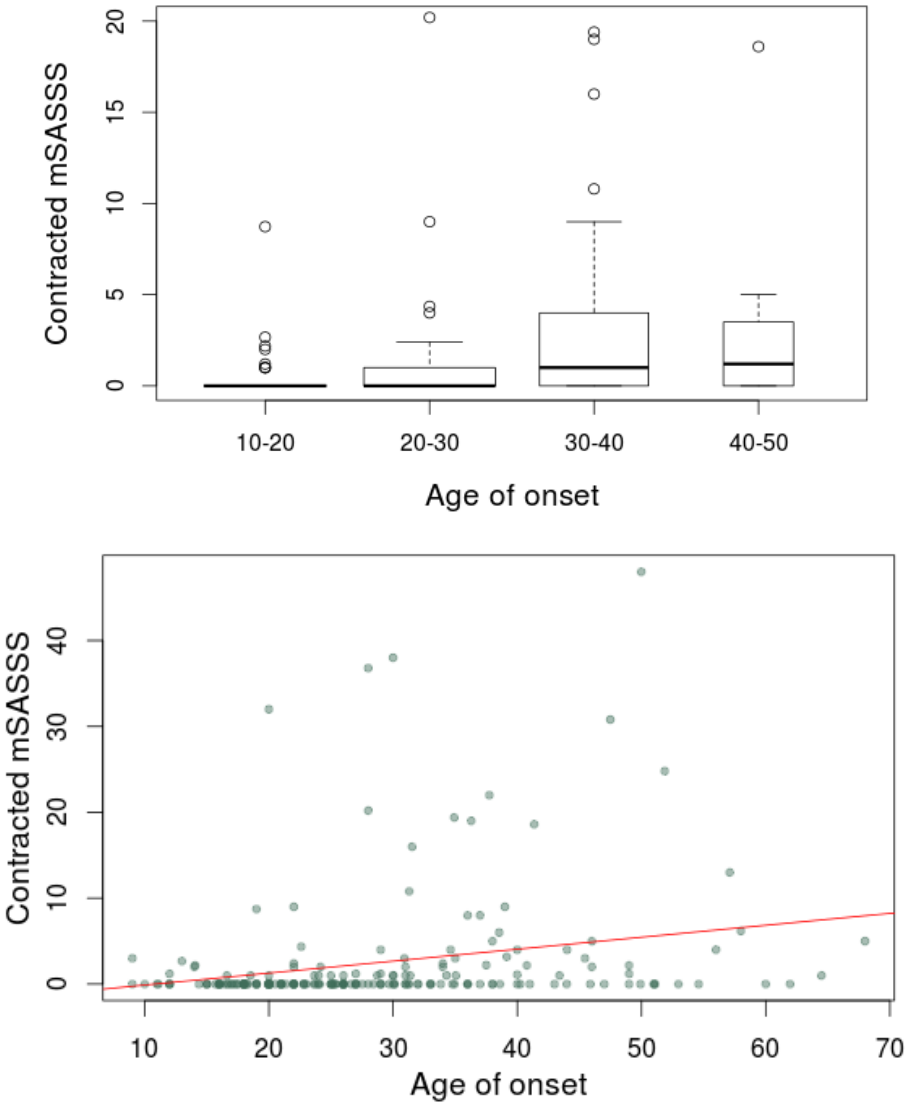


Table 2.11. Comparison of first recorded cmSASSS between age of onset bins. Values are Mann-Whitney test p-values (for comparing non-normal data).

Age of onset group (years)	10-20	20-30	30-40	40-50
10-20				
20-30	0.13			
30-40	3.56×10^{-5}	2.7×10^{-3}		
40-50	1.27×10^{-4}	4.61×10^{-3}	0.69	

This apparent effect of later age of onset causing higher cmSASSS is contributed to by the existence of multiple syndesmophytes in subjects with new onset AS symptoms. Existing syndesmophytes at symptom onset may be caused by minimal symptoms or recall bias. The possibility of this effect being caused by rapidly progressing late-onset AS will be addressed in the longitudinal analysis.

These analyses demonstrate that age of symptom onset is an inconsistent predictor of cmSASSS. Its effect is only present when cmSASSS is corrected for by symptom duration or if symptom duration is limited to more than twenty years. The effect size appears to be small over a lifetime. Because of these inconsistences, and the desire to include subjects with symptom duration less than twenty years in this analysis, age of onset was not used in further multivariable analyses.

Figure 2.16. Age of onset vs latest mSASSS. There is an appearance of lower cmSASSS, corrected for symptom duration, smoking duration, gender and scorer, for subjects with age of symptom onset before sixteen. The difference in mean corrected cmSASSS for figure A is 5.16, significantly different by t test ($p=4.4 \times 10^{-6}$). However use of different sized bins Figure B finds no significantly difference between means (pairwise t tests, $p>0.3$ in each case).

Figure 2.16 A

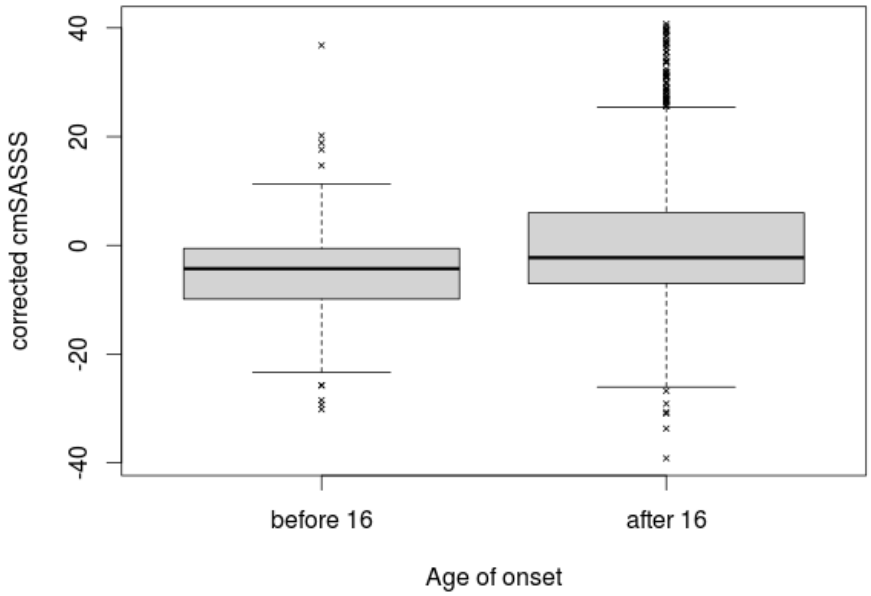
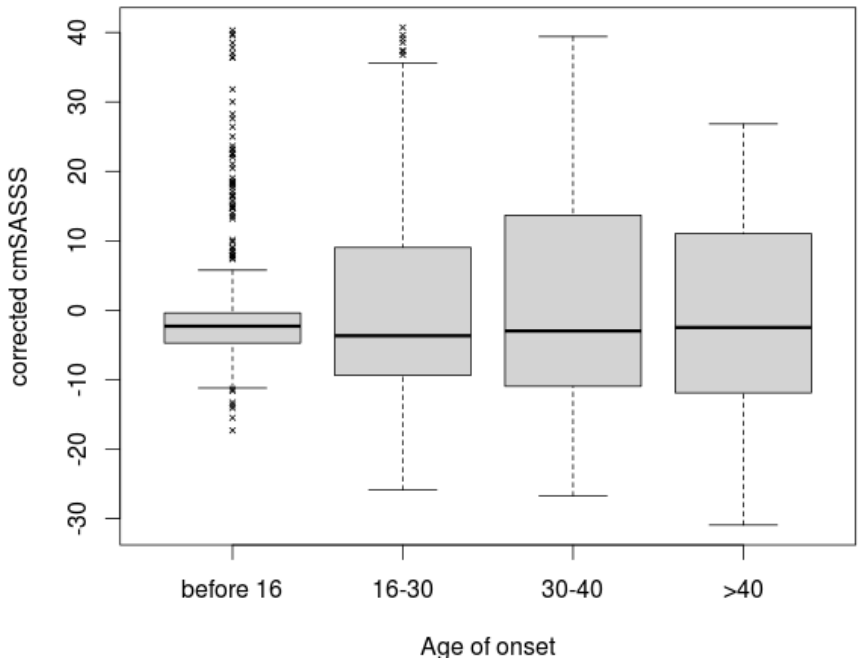


Figure 2.16 B



The best fitting cross-sectional model used fixed effects to correct for symptom duration, male gender and smoking duration (Table 2.12). Adding scorer as a predictor does not improve the fit of the model but would correct any possible scorer based bias, so was included. These models attribute 30-40% of variation in cmSASSS to smoking, symptom duration and gender.

Table 2.12. Model comparison after removal each significant covariate. Model r^2 was estimated by the square of spearman correlation of observed vs predicted values.

Model: cmSASSS vs	Covariate Removed	Estimated model r^2	AIC	LR test vs model A	Wald test vs model A
Symptom duration * smoking duration + Scorer + Gender (Model A)	-	0.389	6473	-	-
Symptom duration + Scorer + Gender (Model B)	Smoking duration	0.355	6516	Inferior $p= 1.65 \times 10^{-10}$	Inferior $p=1.54 \times 10^{-8}$
Symptom duration*smoking duration + Scorer (Model C)	Gender	0.339	6531	Inferior $p<2.2 \times 10^{-16}$	Inferior $p<2.2 \times 10^{-16}$

Table 2.13. Parameter estimates for the count model portion of the final model ('Model A').

Covariate	Estimated Effect Size	Standard Error	P-value
Intercept	1.28	0.20	2.0×10^{-10}
Male Gender	0.58	0.11	1.3×10^{-7}
Scorer 2	0.12	0.16	0.44
Scorer 3	0.36	0.21	0.09
Scorer 4	0.33	0.18	0.22
Scorer 5	0.31	0.21	0.14
Scorer 6	0.54	0.35	0.12
Scorer 7	0.15	0.19	0.42
Symptom Duration	0.03	3.9×10^{-3}	2.0×10^{-12}
Smoking Duration	0.02	7.2×10^{-3}	1.5×10^{-3}
Smoking duration – Symptom duration interaction	3.6×10^{-4}	1.8×10^{-4}	4.5×10^{-2}

2.3.5 Sensitivity Analysis of ZINB Model

Local sensitivity analysis was used to assess the robustness of the model to variation and error in the predictors. 'Model A' from Table 2.12 was chosen as the best model. Notably recall bias could affect symptom duration and smoking duration, and scorers may exhibit bias. The sensitivity analysis, but not multivariable analysis, found the Scorer to be responsible for significant variation in predicted score (Figure 2.16). In the sensitivity analysis, the magnitude of variation is proportionate to the predicted value, and is significantly reduced if scorer 6 is excluded (Table 2.13).

Table 2.14. Sensitivity to changes in Scorer

Symptom duration	Smoking duration	Range of predicted scores: all Scorers	Range of predicted scores: scorers other than scorer 6
0	0	1.69 – 4.83	1.69 – 3.07
0	10	4.25 – 9.52	4.25 - 6.27
40	0	19.4 – 26.6	19.4 – 25.98
40	10	22.95 – 35.5	22.95 – 28.3
40	30	25.01 – 42.96	25.01 – 34.09

A clinically plausible error in symptom duration from recall bias of five years would cause a difference of three cmSASSS points (assuming scorer 2, never smoker, male, symptom duration 20-25 years). A plausible error in smoking duration of ten years would also result in a difference of three cmSASSS points (at symptom duration 0 or 40, male) (Figure 2.16).

Figure 2.17. Predicted cmSASSS over symptom duration. Predicted cmSASSS exhibits inter-scorer variation. Predicted scores for Scorer 6 are higher than the maximum possible cmSASSS of 48. Predicted cmSASSS increases with smoking duration (but smoking duration cannot increase independently of symptom duration).

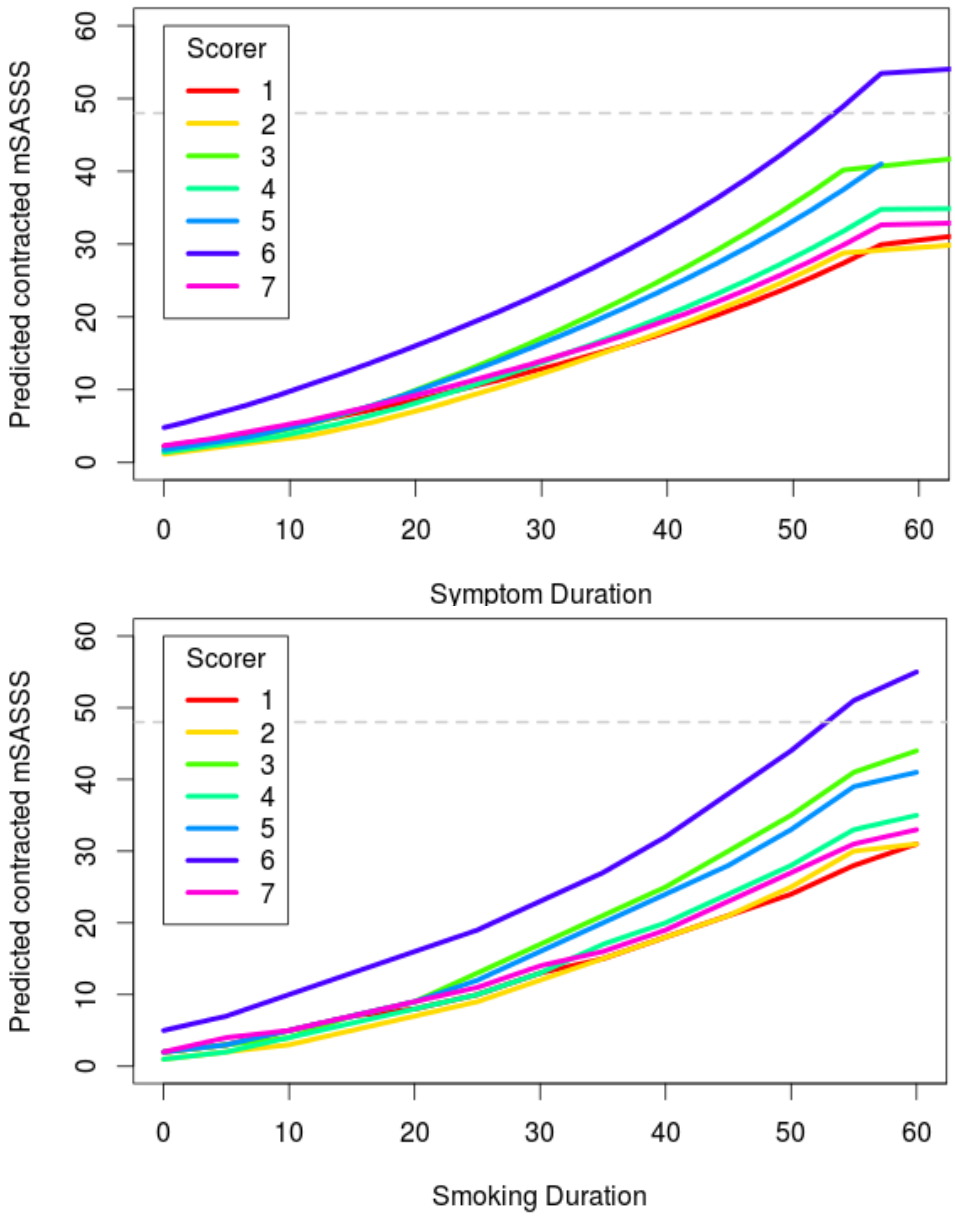
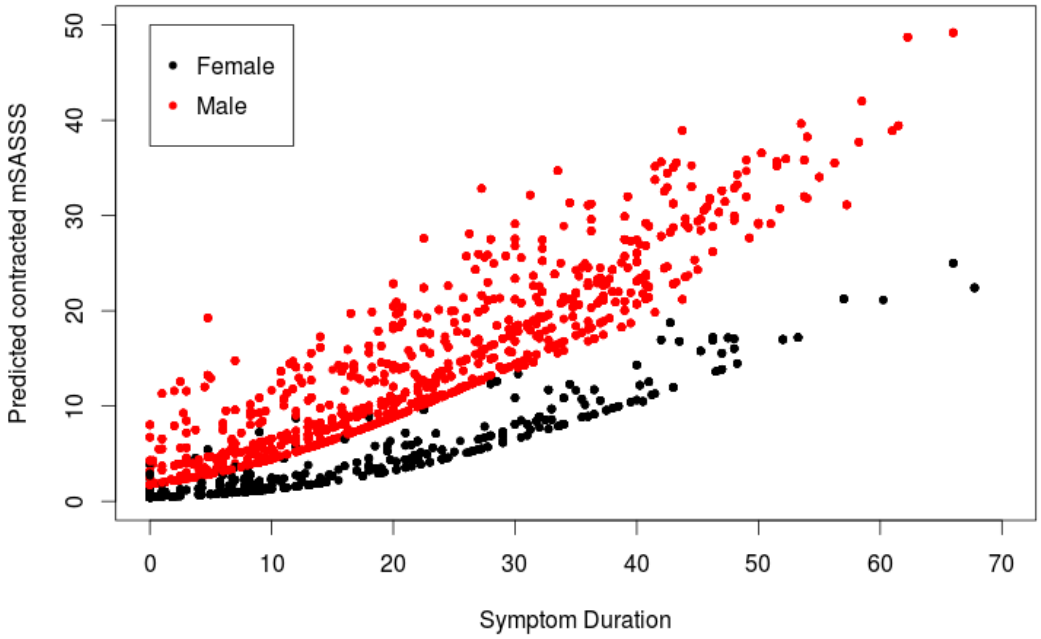


Figure 2.18. Predicted scores by gender. Predictor values used in this plot are real observed values. Predicted zeros can be produced by either the zero model or the count model. In the observed data, females have less severe disease than males for the same symptom duration. Consequently the count model adds a fixed effect for the average influence of male gender (0.6 model units). This gives all males a predicted score slightly higher than zero.



Confidence intervals on effect sizes were refined by simulation.

Table 2.15. Results of confidence interval simulation for count and zero models. As expected, factors such as a long symptom duration, the male gender and a long smoking duration predict higher counts and lower zeros. The confidence intervals around their effect sizes do not cross zero, making them significant predictors according to this analysis. Some of the confidence intervals for Scorers do not cross zero, indicating that they may in fact be significant effects, despite not being considered significant by the multivariable analysis. Given this uncertainty Scorer should remain in the model.

	Count Model		**Zero Model**	
Predictor	**Effect Size Estimate**	**95% Confidence Interval**	**Effect Size Estimate**	**95% Confidence Interval**
Symptom duration	2.85×10^{-2}	2.6×10^{-2}, 3.1×10^{-2}	-0.09	-0.1, -0.09
Male	0.55	0.48, 0.62	-1.07	-1.2, -0.94
Female	Reference value			
Smoking duration	2.34×10^{-2}	1.9×10^{-2}, 2.8×10^{-2}	-0.10	-0.11, -0.09
Scorer 1	-0.12	-0.02, -0.01	-1.01	-1.4, -0.81
Scorer 2	Reference value			
Scorer 3	0.22	0.11, 0.33	0.22	-0.2, 0.45
Scorer 4	0.08	7.8×10^{-3}, 0.16	-0.25	-0.43, -0.07
Scorer 5	0.16	0.06, 0.26	-0.4	-0.75, -0.05
Scorer 6	0.4	0.21, 0.59	-1.1	-1.67, -0.55
Scorer 7 (ST)	0.019	-0.07, 0.1	-0.86	-1.05, -0.65

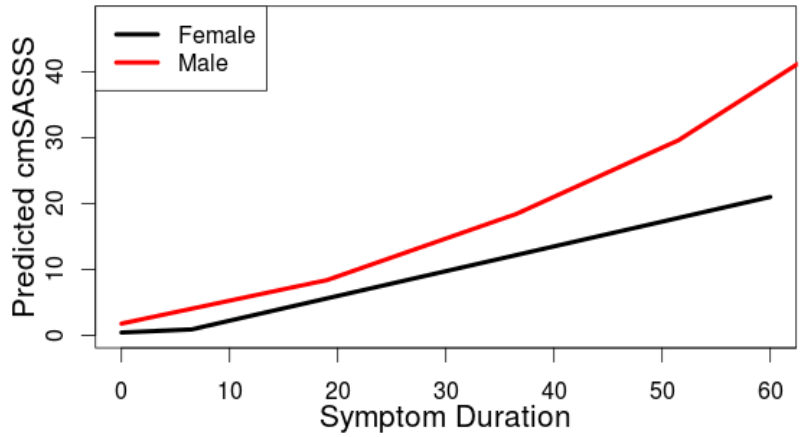
The sensitivity analysis suggested that observations from Scorer 6 may be significantly higher than those of other readers. Multivariable analysis (Table 2.6) did not find any significant effect of Scorer 6. Multivariable analysis compared only Scorer 2 with Scorer 6, and found no significant difference in Scores. Importantly Scorer 6 only scored 15 films used in this study. Sampling bias may have caused an overestimation of Scorer 6's influence in the sensitivity analysis. This small number of films is unlikely to cause bias in a study of almost 1400 films. Importantly all scorers were calibrated, meaning that higher mSASSS from this Scorer may reflect a particular cohort or subset of subjects who Scorer 6 was asked to score. For these reasons films scored by Scorer 6 were not excluded from this study.

Sensitivity analysis is often used to assess the relative contribution of predictors. Comparison of predictor effect sizes from Table 2.14 is inappropriate as the units for gender, symptom duration and smoking duration are not equivalent. To address this issue, the predictive value of each covariate relative to symptom duration can be quantified by performing linear regression of predicted values against the covariate and symptom duration (Figure 2.18) and then comparing the total model coefficient of determination, termed a pseudo-r^2. This analysis was performed using observed values and mirrored Figures 2.15-2.16. It demonstrates that the model attributes the most variation due to symptom duration, then gender, then smoking duration, then scorer.

Table 2.16. Relative importance of covariates. Ranked by pseudo-r^2. Pseudo-r2 = $r^2_{model} - r^2_{null\ model}$

Model	Linear r^2	Pseudo-r^2 relative to symptom duration
Symptom duration only (Null model)	0.71	0
Symptom duration + Gender	0.86	0.16
Symptom duration + Smoking duration	0.82	0.12
Symptom duration + Scorer	0.75 (with or without scorer 6)	0.04

Figure 2.19. Predicted cmSASSS by gender. This null model + Gender was used to calculate pseudo-r^2.



The potential effect of unmeasured confounders was assessed by simulation. For a continuous variable confounder to introduce clinically significant error into the model (eg 1-2 cmSASSS at symptom duration of zero or 5-6 cmSASSS at symptom duration of 20) it would require an effect size of 0.34 model units per standard deviation (Table 2.16).This is approximately half the effect size of symptom duration (if one 'standard deviation' of the non-normal symptom duration is considered to contain 34% of values). For confounding factors that are present or absent, variation in confounder prevalence caused only a few points of cmSASSS variation after twenty years of symptom duration. For a common

confounding factor, with prevalence 43%, an effect size of 0.3-0.43 would be required to cause variation of 4-7 cmSASSS at a symptom duration of twenty years. This is also a large required effect size, more than half that of male gender (0.58) in the same model.

Table 2.17. Smoking metrics by Gender. Males and females have similar rates of 3 level smoking status, with no significant difference upon Chi Squared testing of the 3 proportions. The mean smoking duration between males and females is significantly different by t test ($p=1.3 \times 10^{-6}$).

	Females	**Males**
Never smoker (%)	13.9%	15.9%
Current smoker (%)	25.8%	34.1%
Ex-Smoker (%)	60.3%	50.0%
Mean smoking duration (yrs)	4.4	7.5

Table 2.18. Sensitivity to unmeasured confounders. Predicted cmSASSS assuming male gender, Scorer 2. The continuous confounder is normalised, thus a range of 0-1 is one standard deviation.

Confounder type	Effect Size (ES)	Confounder Prevalence	Symptom Duration	Range of confounder size	Range of predicted cmSASSS	Comment
Continuous	0.34	100%	0	0-1	2.27 – 3.96	Only very large ES cause significant change
Continuous	1	100%	0	0-1	2.27 – 11.33	
Continuous	0.19	100%	20	0-1	9.45 – 12.47	An ES of 0.34 causes cmSASSS to vary by 5
Continuous	0.34	100%	20	0-1	9.45 – 14.41	
Continuous	0.70	100%	20	0-1	9.45 – 21.65	
Present / Absent	0.18	25%	0	Present or Absent	1.90 – 2.70	ES or prevalance variation has little effect at this symptom duration
Present / Absent	0.49	60%	0		1.29 – 3.23	
Present / Absent	0.88	25%	0		0.89-2.89	
Present / Absent	0.43	27%	20		7.16 - 14.37	Prevalence variation causes minor change
Present / Absent	0.43	72%	20		5.26 - 11.12	
Present / Absent	0.3	43%	20		7.34 - 11.57	ES > 0.3 required for significant change
Present / Absent	0.6	75%	20		4.20 – 10.75	

2.4 Discussion

The TASC cohort is the largest AS cohort ever investigated for radiographic severity. Its demographics and incidence of AS extra-articular manifestations are similar to other cohorts. Uveitis is common and its cumulative incidence increases with symptom duration. Most subjects are ex-smokers who report a minimal smoking duration.

Inter-rater reliability was high, likely aided by the removal of scores for erosion, sclerosis and squaring – the more subjective features of mSASSS. Univariable, multivariable and sensitivity analysis all suggested that cmSASSS from particular scorers were higher than the reference scorer, Scorer 2. This may be due to sampling error, more severe cohort of subjects, due to clinic or location related confounders, site specific variation in radiology equipment, or bias in cmSASSS scoring. Sensitivity analysis indicated that the model predicts higher cmSASSS from Scorer 6, likely a result of sampling error as only 15 cmSASSS from Scorer 6 were used. Apart from a possible effect of Scorer 6, the model appears robust, or not overly sensitive, to plausible errors in input data. Reasonable events such as a recall bias of 5 years for symptom duration would only result in a 3 point difference in predicted cmSASSS.

The distribution of cmSASSS was zero-inflated. The disproportionate number of subjects with a low symptom duration would have contributed to this, but many subjects with symptom duration > 30 had cmSASSS of zero. The left-skewed, overdispersed, zero-inflated distribution of cmSASSS greatly complicated statistical modelling. Inclusion of cmSASSS of zero is important, as subjects with mild disease need to be compared for protective features. Modelling of this cross-sectional data attributed 30-40% of variation in cmSASSS to gender, symptom duration, smoking duration, following correction for inter-rater variation. By this measure, this model is superior to previous efforts at modelling mSASSS.

Cervical syndesmophytes were more common and appeared earlier than lumbar syndesmophytes in both genders. Cervical ankylosis was far more common than lumbar

ankylosis. The severity of cervical radiographic disease contrasts classical descriptions of AS, as reflected by the modified New York Criteria, which consider inflammatory lumbar pain but not cervical pain to be a sign of AS. The reason why cervical radiographic disease is more severe than lumbar is not known, but may be related to biomechanical differences such as the higher mobility of the cervical spine, or possibly differential gene expression in bone or cartilage across the spine.

The association between later age of onset and more syndesmophytes has been seen in previous studies (71, 105). In this cohort the association was not consistently present. In particular, models including all symptom durations and models that corrected for age instead of symptom duration found no effect of age of onset. The association is contributed to by late onset AS subjects having multiple syndesmophytes present within five years of symptom onset. This could be from recall bias, or may be because some subjects initially suffer minimal symptoms. Nevertheless it is a sign that symptom duration can be an unreliable measure of AS disease duration, making symptom duration lower than true disease duration. The presence of self-reported age of onset after 60 years in some subjects supports this. This analysis found no consistent evidence of any effect of age of symptom onset, with any possible effect size being very small and only evident after twenty years of symptoms. To answer the question of whether juvenile symptom onset is protective, future models should restrict symptom duration to over twenty years, separate age of onset into bins to isolate any protective effect of younger onset from the anomalous association of late onset symptoms with higher cmSASSS, and perform tests correcting for symptom duration and current age.

Consistent with previous data, syndesmophyte formation increases with symptom duration, age and smoking dose. Male gender predisposes to more severe disease at all points of symptom duration, with equal effects early and late in the disease. Smoking is a significant predictor of cmSASSS in a dose dependent manner. Smoking duration was more predictive than pack year history. This comparison of pack year history to smoking duration has not been examined by previous publications. The difference between the two measures may be due to less recall bias of smoking duration compared to cigarettes per

day, or could be caused by the stronger effect of smoking exposure than high smoking dose. Smoking is believed to promote osteoproliferation by promoting systemic inflammation (82).

Gender was seen to influence severity in a number of ways. Males had more severe disease at all points of symptom duration, and were more likely to progress to ankylosis. In females, cervical ankylosis was uncommon and lumbar ankylosis rare. The effect of gender on radiographic severity was greater than that of smoking. Males were more likely to have lumbar or cervical syndesmophytes, and were more likely to have cervical predominant disease than females. The prevalence of cervical predominant disease in females was affected by the high frequency of scores of zero in females. Cervical predominant disease was more common than lumbar predominant disease, especially at low symptom duration. Epidemiological studies that identify gender as a risk factor must consider the possibility that gender is an indirect risk factor with all or part of its effect mediated by a true biological determinant. In this case, gender could act as an indirect marker of smoking status or occupational status, which are both associated with radiographic severity. Gender is unlikely to be a marker for smoking status in this study, as 74% of subjects are male, the model corrects for smoking status, and the contribution of gender is greater than that of smoking. In the 3 level smoking model, gender should not act as an indirect marker of smoking as there is no significant difference in rates of 3 level smoking between males and females. The effect of gender may contain some occupation related effect, but this is likely to be limited – the effect of gender is already evident in subjects with low symptom durations. This translates to early in subjects' working lives as the mean age of onset is 25 years. Gender is the most influential predictor of radiographic severity identified in this study, with particularly strong effects on lumbar syndesmophytes. Even in the cervical spine, for a given symptom duration the average male has more than twice as many syndesmophytes. Gender related genetic and hormonal differences should be investigated to determine their potent disease influencing mechanisms.

Some covariates are difficult to use as predictors in cross-sectional analyses, especially those that fluctuate regularly, as they may have only been recorded as a single

observation at one point in time. This applies to markers of disease activity such as ESR and CRP, symptom scores such as BASDAI and VAS stiffness, and quantification of weekly exercise. Usual analysis of disease activity markers involves quantifying their area-under-the curve values which is used as a predictor for longitudinal analysis. Tests such as CRP also have large inter-institution variation due to different assay preparations. Symptom based scores, symptom duration and smoking duration can all exhibit recall bias. Recall bias may also be present in data on extra-axial manifestations of AS, as (for the TASC and SPARRC cohorts) these results contain a mixture of physician-diagnosed and self-reported uveitis, IBD and psoriasis. Functional scores are both markers and predictors of severe disease. Many are highly correlated with mSASSS, and in this analysis, outcome measures including BASFI were highly significant predictors of cmSASSS.

Any significant association found between disease activity measures and radiographic severity must be interpreted with caution. Measures of disease activity such as BASDAI are a phenotype or outcome, influenced by many factors including medications, physical factors such as occupation and exercise, and may be correlated with severe radiographic disease.

Longitudinal analysis, which compares change in these scores to change in mSASSS, is the best method of demonstrating the significance of these covariates. A recent cross-sectional analysis of the OASIS cohort after twelve years did not find BASDAI to be predictive of mSASSS, but using a longitudinal analysis, the same data demonstrated that BASDAI was predictive of radiographic progression, after a two year time lag (93). Similarly, covariates such as recent stiffness and number of back exercises per week are best analysed in a longitudinal model, and were statistically not significant in the TASC cross-sectional model.

Covariates in cross-sectional model must be provided to the model in a metric appropriate to the biological context and the features of that covariate. Smoking is best measured by a cumulative metric such as pack year history or smoking duration if it is hypothesised that

cumulative smoking dose influences the outcome. Covariates that fluctuate dramatically over time should also be considered using a cumulative measure such as the area-under-the-curve over time. Covariates that do not change such as gender or family history should not have a cumulative metric unless an interaction with time has been established. For example uveitis has a cumulative incidence, and a lifetime history of no episodes of uveitis in an 18 year old is not the same as no episodes of uveitis in an 80 year old.

2.5 Conclusion

This study describes and quantifies patterns and predictors of cross sectional radiographic severity. The previously identified predictors smoking, gender and symptom duration were found to explain about 40% of variation in radiographic severity. Smoking duration, which has not previously been compared to pack year history, was found to be a more accurate predictor than pack year history. Other covariates found to be significant by previous studies were either not recorded or not recorded in a form appropriate for cross-sectional analysis. Syndesmophyte formation in males was far more severe, especially in the lumbar spine. Even after attributing variation to gender, symptom duration and smoking duration, variation in radiographic severity is large. The source of this variation as well as the mechanisms of gender based variation should be investigated to predict and prevent osteoproliferation.

3. Genetic Investigations

3.1 Introduction

Radiographic severity in AS is variable and highly correlated with functional ability. One quarter to one third of variation in radiographic severity has previously been attributed to non-genetic factors. The remaining variation remains unexplained. A large proportion of variation in AS susceptibility is genetic and pedigree studies have found strong evidence of heritability of severity. Consequently it is believed that much of the unattributed variation in radiographic severity is determined by genetic factors. To date, a number of candidate gene studies have investigated suspected loci, but no results have been reproducible. Importantly researchers have specifically chosen these candidate genes for investigation.

This study is the first genome wide search for determinants of this phenotype. The extensive cost and logistics of collecting radiographs, clinical determinants of radiographic severity and genotypes from thousands of cases has been an enormous challenge in this area. Description and analysis of this radiographic phenotype requires complex non-parametric statistics that were developed in the 1990s. The statistical genetic and genotyping tools used in this study were developed in the last ten years. As a result this first GWAS of radiographic severity in AS was challenging in many respects. This study and subsequent studies should provide insights into AS pathophysiology that will improve prognostication and treatment.

3.2 Methods

3.2.1 Collection and Analysis of Phenotypic Data

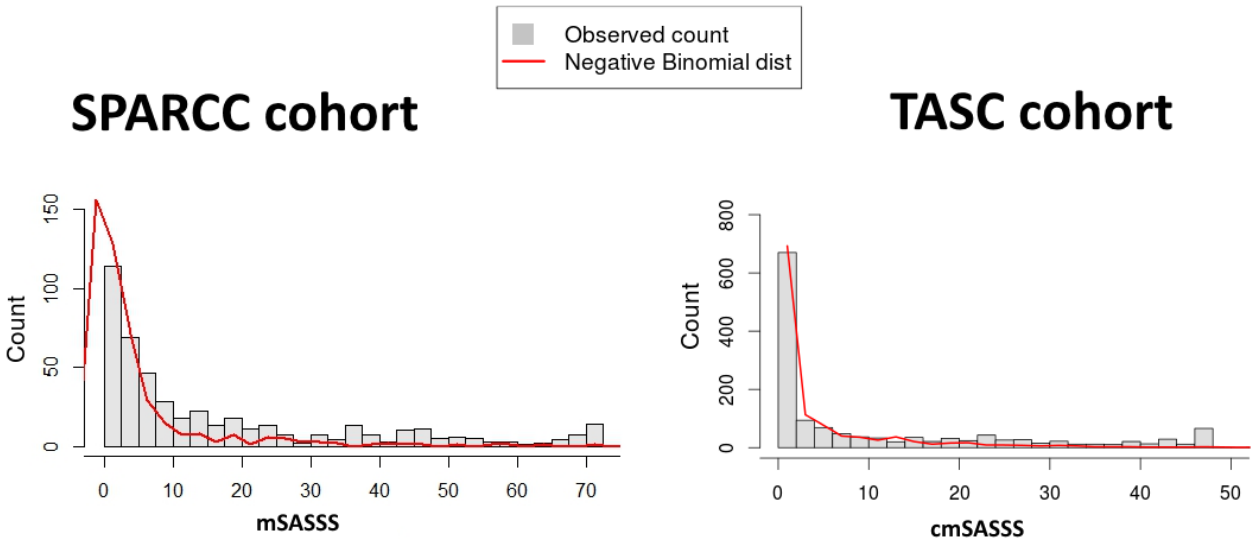
Radiographs and clinical covariates of >1000 TASC subjects were collected as outlined in Chapter 2. The Spondyloarthritis Research Consortium of Canada (SPARCC) collected mSASSS and clinical covariates on 274 Canadian subjects used in this study, with local ethics committee approval. SPARCC subjects did not have per corner mSASSS available, thus scores for erosion, sclerosis and squaring could not be removed. Scores from SPARCC subjects were described as total mSASSS, cervical mSASSS and lumbar mSASSS. Clinical covariates were only available for subjects from two centres. 168 SPARCC subjects from one centre had smoking duration and pack year history while the remaining 106 subjects from the other centre had three level smoking status (current/ex/never) but not smoking duration. All SPARCC subjects had gender and symptom duration, but data on other covariates were not available.

The distribution of SPARCC mSASSS resembled that of TASC cmSASSS and the negative binomial distribution (Figure 3.1). Modelling was performed using NB and ZINB models, as described in Chapter 2. The ZINB model fit was superior as demonstrated by Vuong's test, AIC and standard goodness-of-fit plots (Appendix A.5). The TASC and SPARCC cohorts were modelled separately due to the use of cmSASSS and mSASSS respectively. For the GWAS, the two datasets were only combined at the stage of GWAS meta-analysis, to use data using two different measures of the same trait.

For the TASC cohort, the scoring sheets for mSASSS were available. This allowed scores of 1 to be removed for conversion to cmSASSS. As previously discussed, for reasons such as biological relevance, the cmSASSS score is preferable to mSASSS. Unfortunately the original scoring sheets for the SPARCC cohort were not available, and these scores were not able to be converted to cmSASSS. For this reason association testing was performed on SPARCC scores separately, then meta-analysed with the TASC association testing results. The two scoring systems, cmSASSS and mSASSS both describe the AS

phenotype, and scores of 1 are uncommon – they contribute to <5% of changes over a two year period (14, 55, 56). This study performs independent association testing of each cohort to assesses the significance of each SNP. Thus main role of the phenotype metric is to compare the severity of each subject with this sample of the entire AS population. The per SNP significance results can be meta-analysed so long as they are treated as different cohorts by the statistical analysis. Even when the same phenotype metric is used, association testing is often performed on each cohort separately then meta-analysed, which allows for variation between the cohorts caused by unmeasured effects and known cohort specific covariates (136).

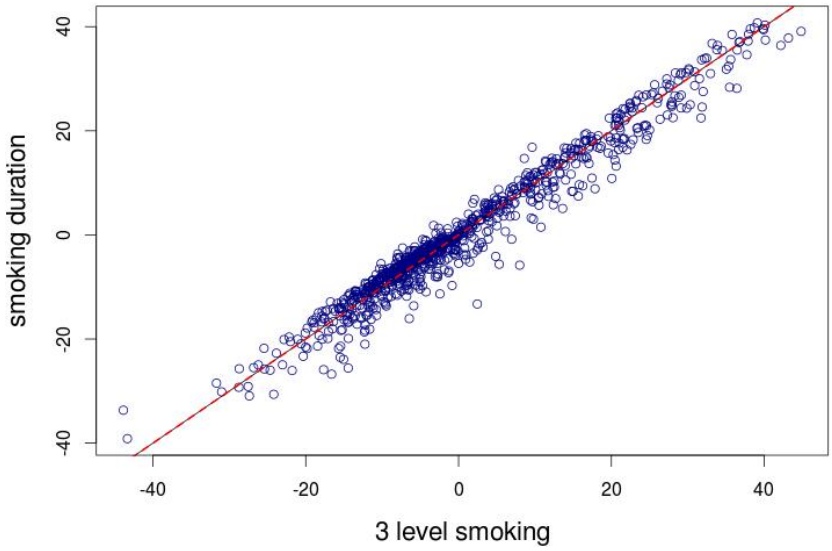
Figure 3.1. Radiographic scores in each cohort. The SPARCC mSASSS scores can be approximated to the negative binomial distribution. The negative binomial distribution assumes higher scores are less likely. NB and ZINB models will therefore predict lower values for subjects with the most severe disease. The sample size of the TASC cohort is much larger than the SPARCC cohort with 979 and 382 subjects respectively. Both distributions have a predominance of low scores.



To maximise subject numbers in the GWAS, it was desirable to include subjects with three level smoking status but not smoking duration data. Models using three level smoking status and smoking duration were previously compared by AIC, likelihood ratio tests and

Wald tests and were found to be almost the same as the smoking duration model (Table 2.8). Analysis comparing the three level smoking model with the smoking duration model found a Spearman correlation 0.97, and ICC 0.98 (95% CI 0.977-0.982). As the residuals of these two models were almost identical, for subsequent steps the three level smoking status model was used.

Figure 3.2. Comparison of residuals from different ZINB smoking models. TASC cohort data. There is little variation between the two models, as demonstrated by the proximity to perfect correlation (r = 1).



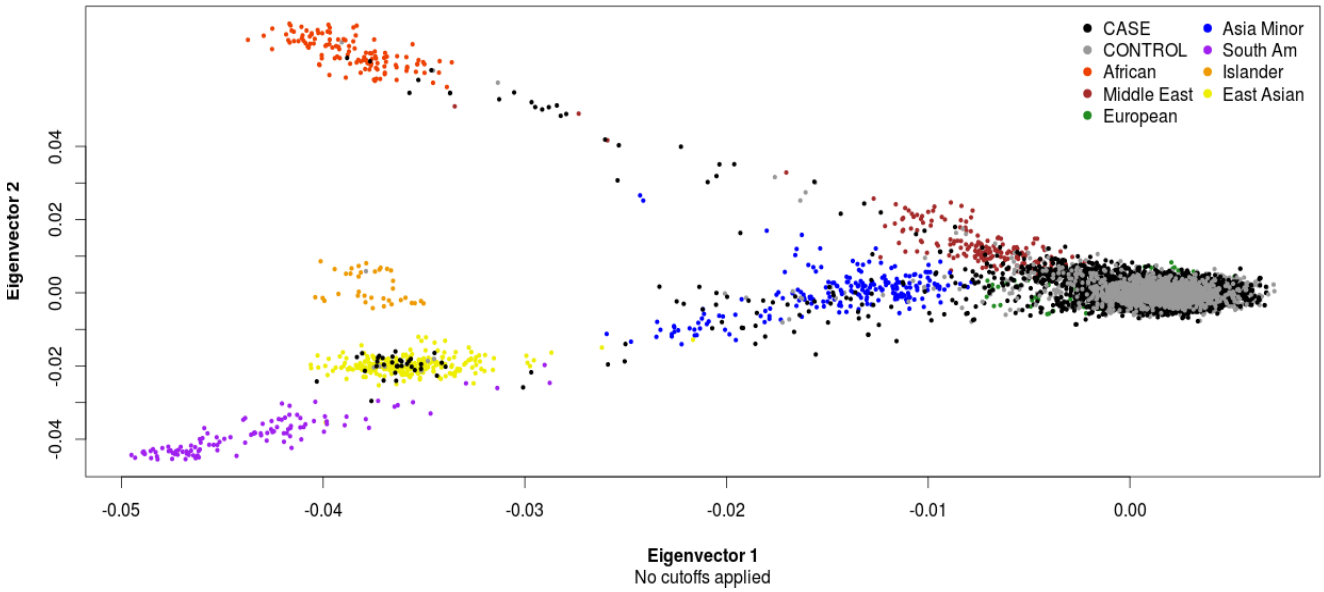
3.2.2 Genotyping and Data Preparation

SNP genotyping was previously performed during multiple studies of AS susceptibility, involving subjects from TASC and SPARCC (22, 35, 36). TASC subjects were genotyped on at least one of the Illumina HumHap370, Illumina Human660W-Quad or Illumina Infinium Immunochip platforms. SPARCC subjects were genotyped on at least one of the Illumina Human OmniExpress or Illumina Infinium Immunochip studies.

Genetic quality control of Immunochip genotypes using PLINK v1.07 (137) was performed by Adrian Cortes (AC), and for all other genotypes by ST. Only autosomal SNPs were studied. Genotypes from Immunochip control subjects were used for quality control of Immunochip genotypes, and genotypes from Wellcome Trust Case Control Consortium (WTCCC) controls for Illumina HumHap370 and Illumina Human660W-Quad data. Illumina Human Omniexpress genotypes underwent quality control using cases and controls genotyped on this chip. Control data was used for heterozygosity and Hardy-Weinberg Equilibrium testing. SNPs with genotype call rates <99% were excluded. Individuals were excluded if their heterozygosity was greater than 2 standard deviations from the mean or if >3% of SNPs were missing. Rare SNPs with Minor Allele Frequency (MAF) < 5% in cases were excluded. SNPs not in Hardy-Weinberg equilibrium ($p<1x10^{-5}$) were removed.

AC assessed population stratification of Immunochip genotypes using 5 Principal Components (PCs) and removed non-Caucasian outliers. ST combined these genotypes with the genotypes from the other SNP chips and recalculated PCs using the s*hellfish* software package. *Shellfish* has been validated against the commonly used PCA software and has been used in large GWAS (138, 139). Non-Caucasian outliers were excluded if either of their first two PCs were greater than six standard deviations from the mean. A six standard deviation cut-off is commonly used and is employed by EIGENSTRAT (140). Non-Caucasians were excluded because population stratification could not be reliably corrected across different ethnicities.

Figure 3.3. PC analysis displaying the first two eigenvectors. Colours represent Hapmap ethnic groups. Numerous cases (black circles) were clearly not of Causasian ethnicity.



Identity By State calculation using PLINK (PI_HAT > 0.9) was used to identify duplicate samples and to match subjects who had been genotyped on multiple SNP chips.

In total, QC removed 82 subjects from the TASC GWAS studies and 90 from the SPARCC GWAS study. Twenty-one subjects from the Immunochip study who AC performed QC on were removed by the PC analysis. The final counts of subjects with sufficient data for association testing are presented in Table 3.1.

Table 3.1. Subjects counts for association testing. These subjects all have a genotype that has passed quality control, smoking data, gender, symptom duration and mSASSS.

Cohort	GWAS chip* only	Immunochip only	GWAS + Immunochip	Total
TASC	108	252	619	979
SPARCC	75	21	286	382

* GWAS chips: Illumina HumHap370, Illumina Human660W-Quad, Illumina Human OmniExpress.

Table 3.2. Imputed SNPs available per subject.

Cohort	GWAS chip only*	Immunochip only	GWAS + Immunochip
TASC	5-6 x 10^6	500 000	5-6 x 10^6
SPARCC	7 x 10^6		7 x 10^6

Haplotype phasing and imputation of nearby loci was performed using SHAPEIT and IMPUTE2 respectively. A merged reference sequence was used for imputation, consisting of 1000 Genomes phase 3 and UK10K. Validation studies of the sequential SHAPEIT-IMPUTE2 method (141) found that when using the 1000 Genomes reference panel, SNPs were imputed with high accuracy (Table 3.3).

Table 3.3. Imputation accuracy of SHAPEIT-IMPUTE2 (141). Mean R^2 refers to the coefficient of determination between a test panel of genotyped SNPs and SNPs imputed using data with the genotyped SNPs missing ('masked genotyped SNPs').

	MAF 0.01 to 0.03	MAF 0.03 to 0.05	MAF > 0.05
Mean r^2	0.82	0.86	0.92

Only high confidence imputed SNPs were included, by removing imputed SNPs with an imputation score <=0.8. Imputed SNPs with a MAF <0.05 were removed due to the limited sample size and therefore power. For example, a SNP with MAF 0.02 genotyped in 1000 subjects would only be present in 20 subjects. Comparing these 20 phenotypes with 980 would be prone to error. Following imputation, subjects genotyped on multiple SNP chips had their data merged into one subject entry.

Imputed variants with MAF < 0.05 are often excluded from analysis due to significant uncertainty during imputation (142) and association testing (143). SNPs with a population

MAF < 0.05 are more sensitive to outliers and have less reliable imputation, which is reflected by lower average imputation r^2 values (144).

Imputed 'best-guess' genotypes are not perfect, so measures must be taken to minimize genotype misspecification. Common techniques described by De Bakker *et al.* include only using SNPs genotyped on at least 1 platform and removal of SNPs with poor expected imputation performance (145). SNPs with poor expected imputation performance are those with a low imputation quality and those with MAF <0.05. SNPs with a low imputation quality score are those that possess low LD with tag SNP. These SNPs are not imputed with high certainty and clearly should be excluded. Once MAF falls below 0.05, imputation accuracy (using Hapmap Phase II reference panel) falls dramatically (136). This finding appears to occur even after the application of an imputation quality cutoff. Howie *et al*. found a similar fall in accuracy when MAF falls below 0.05 when performing pre-phasing with SHAPEIT then imputation with IMPUTE2 using the 1000 Genomes reference panel (141).

SNPs with low MAF may have imputation (confidence) scores that appear higher than those of common variants (143). Most rare SNPs will be homozygotes for the common allele, making the genotype easier to predict which falsely elevating the r^2. Asimit and Zeggini suggest that alternative measures of imputation accuracy are required for rare SNPs to avoid genotype misspecification and an increase in false positive association (143).

This study of AS radiographic severity imputed genotypes of AS subjects using a combined reference panel from 1000 Genomes and UK10K, which should include greater haplotype diversity than Hapmap phase II and 1000 Genomes. Nevertheless, the current study involves less than 1500 subjects in total, and many SNPs were imputed in <1000 subjects. This means that a SNP with MAF 0.03 which was imputed in 800 subjects would have the minor allele in only 24 subjects. Any significant outliers or sampling bias in this population could create a false positive result. Therefore in this case, the stringent

principles established by De Bakker *et al*., especially the exclusion of rare imputed SNPS, should be used in this study. Additionally this analysis used the standard IMPUTE2 protocol which treats imputed SNPs the same as genotyped SNPs during the association analysis. All imputed SNPs need to be imputed with high confidence. This concern is addressed by other imputation protocols, such as the MACH imputation package, by considering the uncertainty of imputed genotypes during association testing (using their r^2).

3.2.3 *A priori* Power Calculation

GWAS power was calculated using the quantitative trait disequilibrium test (146) employed by the Genetic Power Calculator (147). This commonly used method compares the distribution of the test statistic under the null hypothesis with that under the alternate hypothesis and can be used for quantitative phenotypes that are not grossly non-normal (148). Power to detect a range of MAFs and effect sizes were calculated for a scaled Linkage Disequilibrium coefficient (D') of 0.8.

A D' value of 0.8 between tag SNPs and causal variants is a commonly used cutoff of significance (149). The cutoff point is an arbitrary degree of linkage disequilibrium which allows for minimal loss of power. As D' approaches 1 there no loss of power in the assumption that the tag SNP perfectly tags the causal variant (144, 150). Use of a higher D' cutoff would decrease the number of imputed SNPs available for analysis. In this case a D' cutoff of 0.9 would decrease the number of imputed SNPs by 24%. A lower D' cutoff (eg 0.6) would increase the uncertainty of imputation which would increase the number of false negatives and positives upon association testing.

The Genetic Power Calculator can be used for studies where the distribution of the test statistic is known under the null and alternate hypotheses (148). Tests statistics in genetic association studies most commonly follow a normal or chi squared distribution. This finding is consistent with the central limit theorem which states that a test statistic that consists of additive contributions from multiple observations will tend to have a normal distribution when the sample size is large (151). To apply this principle to this study we should examine the expected test statistic. As the genetic association test for this GWAS is a linear model, so long as the phenotype is not grossly non-normal, the expected test statistic from millions of SNPs should be normal. The gross approximation the phenotype to a normal distribution is evident in Figure 3.4. Post-hoc examination of the test statistic confirms that it is in fact normally distributed (Figure 3.5).

Figure 3.4. Distribution of residuals from the 3 level smoking model. The distribution is grossly bell shaped.

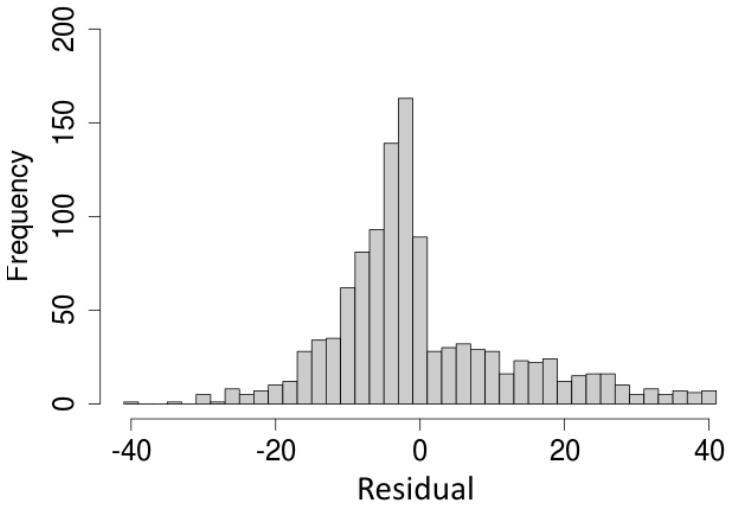
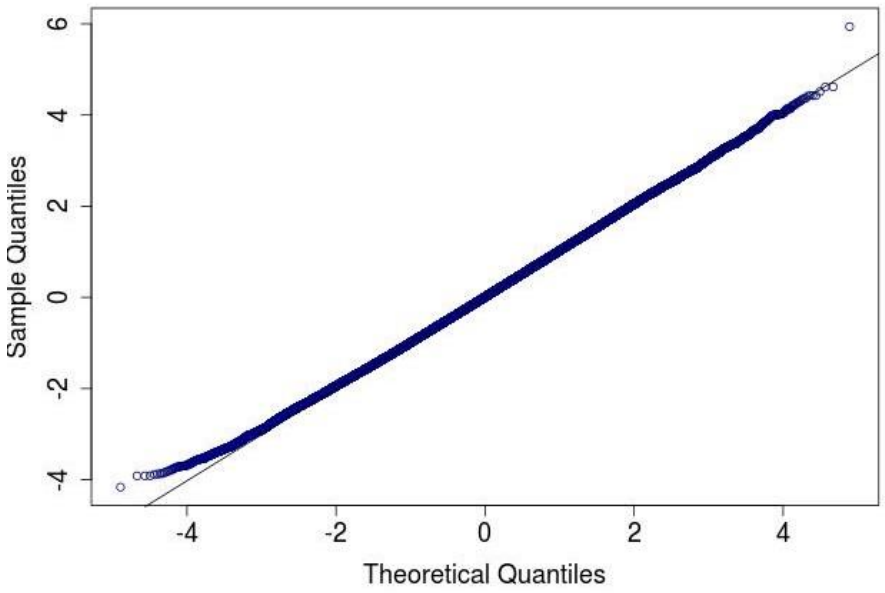


Figure 3.5 QQ plot of association test statistic from the SPARCC cohort. As predicted by the central limit theorem, the test statistic is normally distributed, indicating that use of the Genetic Power Calculator was appropriate.



3.2.4 Genome Wide Association Testing

Residuals from the three level smoking TASC and SPARCC ZINB models discussed in Chapters 2 and 3.2.1 were used as quantitative phenotypes for association testing. For the TASC cohort, these residuals represent cmSASSS corrected for symptom duration, gender and three level smoking status. For the SPARCC cohort, the residuals represent mSASSS corrected for the same covariates. This phenotype represents radiographic severity within the population of AS subjects, so disease free controls cannot be assigned a phenotypic score.

GWA testing was conducted to identify individual loci associated with severe disease. These loci, for example transcripts or regulatory elements, could in future be compared in vitro between AS subjects and healthy controls. Unlike candidate gene studies, GWAS does not only examine loci hypothesised as being relevant. Early candidate gene studies have also been criticized for lack of correction for population stratification and for inappropriate p-value thresholds (152). GWAS is most able to detect common variants with large effect sizes, whereas candidate gene studies are often designed to detect the effects of rare variants.

Association testing between the 3 level smoking model residuals and imputed genotypes was performed in PLINK separately for each cohort. The first 3 PCs were used as covariates to control for subtle population structure, and the residuals were treated as a quantitative phenotype. This study used an additive genetic model performed by linear regression. The phenotype was regressed against the genotype of each SNP (homozygous minor, heterozygous, homozygous major). This analysis treats all SNPs as biallelic, which is a common assumption in GWAS (152).

As per standard IMPUTE2 protocols, during association testing high quality imputed SNPs ('info score' >=0.8) were treated as if they were genotyped with 100% certainty (153). Other imputation protocols weight imputed SNPs by their confidence score (info score) to reflect uncertainty in their genotype, and weight genotyped SNPs as 1.

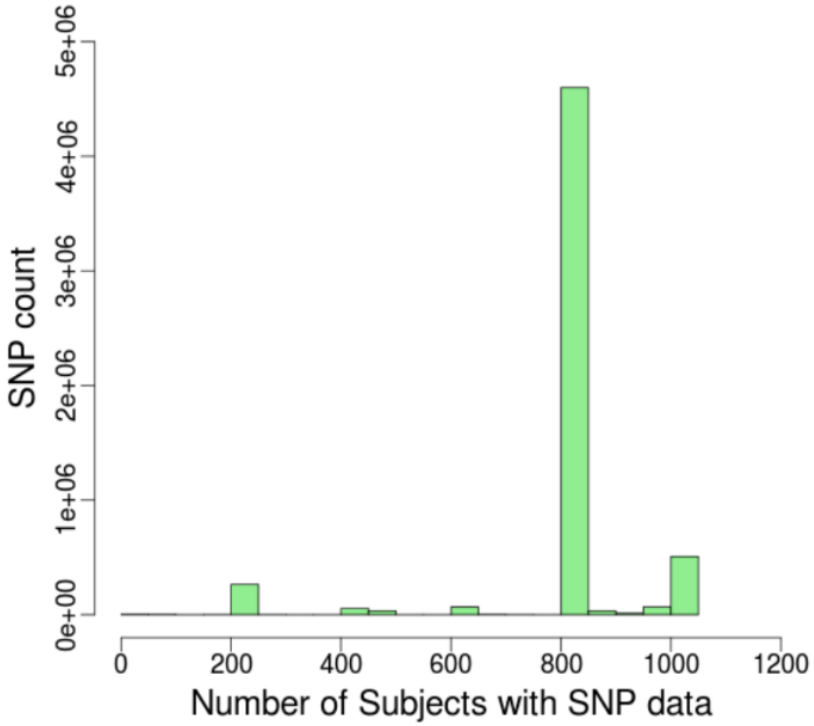
Association testing of genome wide imputed alleles was performed in PLINK using 3 PCs as covariates to control for subtle population structure. The high confidence imputed SNPs were weighted equally to genotyped SNPs during association testing. As the TASC cohort had cmSASSS values, while SPARCC had mSASSS values, the two datasets were analysed as two separate GWAS. The Genomic Inflation Factor (λ) was calculated by PLINK.

Multiple methods of GWAS meta-analysis have been developed. The most common methods are: fixed effect meta-analysis, random effect meta-analysis and p-value meta-analysis (also known as "z-score meta-analysis"). Fixed effect meta-analysis assumes that the true effect of each risk allele is the same in each cohort, while random effect analysis does not (142). P-value meta-analysis creates a z-score for the association for each SNP, representing the magnitude and direction of effect. The score is compared to the average z-score for all SNPs, then weighted by the sample size of each cohort, and results from both cohorts are combined. The comparison of each SNP's effect to the average effect allows each cohort to have a unique scale of measurement (eg cmSASSS and mSASSS) (154). Unfortunately this method provides no estimate of effect size for any SNP, only effect direction. The main benefit of p-value meta-analysis in this case is its ability to compare two cohorts using two different metrics for the same phenotype. This is possible because it only compares the z-scores (which are standardised), not the effect size (which is scaled in the dimensions of the phenotype).

P-value meta-analysis of the two datasets was performed using METAL (155). METAL meta-analyses converts the direction of effect and observed p-value from each cohort into a signed z-score, and weights the score by the number of observations from each cohort. This method provides a meta-analysed p-value, but does not provide an effect size. Other meta-analysis techniques would be inappropriate as they assume the effect size from each cohort uses the same scale. Instead standardised effect size from the TASC cohort are reported. Meta-analysis results were visualised by Manhattan plot and LocusZoom regional association plots (156). Risk allele frequencies were reported from the 1000 Genomes phase 3 dataset.

Gene, exon and expression data were collected from NCBI gene release 106 (157) and ENSEMBL build 78 (158) and FANTOM5 (159). Structural and regulatory features were visualised using the UCSC genome browser (160).

Figure 3.6. Histogram of SNP count per subject after meta-analysis. The vast majority of SNPs were imputed in at least 800 subjects. These subjects also had data for symptom duration, smoking status, gender and mSASSS or cmSASSS.



3.2.5 HLA Imputation

The HLA region is a unique and complex portion of the human genome. Imputation of this region is best performed with specific techniques that impute HLA alleles using the LD patterns of multiple surrounding SNPs (161).

Accurate genotyping of this region was considered important in this study due to the strong association between HLA-B27 and AS susceptibility. Previous studies have suggested that HLA alleles other than HLA-B27 may be associated with bone formation in AS (71).

The software SNP2HLA imputes SNPs in the HLA region and reports which classical HLA alleles and amino acids they encode. SNP2HLA uses the Beagle software package (162) and a database of reference HLA genotypes from the WTCCC. SNP imputation assumes that the haplotype distribution in cases and controls matches that of the reference population. The most likely haplotype and genotype for each individual is modeled using methods specific to each software package (eg a Hidden Markov Model or a graphical model).

Human Leukocyte Antigen (HLA) association testing was performed as multiple HLA alleles have been implicated in AS susceptibility. Imputation of HLA alleles was performed using the SNP2HLA software package using a Type 1 Diabetes reference panel (163). SNP2HLA imputes classical HLA Class I and II alleles, as well as SNPs across the Major Histocompatibility Complex and where within HLA genes, the amino acids they encode. It is used for high throughput HLA imputation and has successfully identified risk HLA alleles in IBD, AS, and lymphoma (164, 165).

SNP2HLA imputation was only performed using Illumina Immunochip genotypes, from 1144 subjects. Genotyping and quality control was performed as detailed in Section 3.2.2.

Alleles with SNP2HLA imputation r^2 <0.6 were excluded, leaving 7674 alleles (note this r^2 score is a different metric to the info score used by IMPUTE2). The average accuracy of SNP2HLA imputation when using Immunochip genotypes and the Type 1 Diabetes reference panel was 96.7% for HLA alleles (4 digit resolution) and 99.3% for amino acid polymorphisms. Association testing using PLINK was performed separately for the TASC and SPARCC cohorts. Residuals from the 3 level smoking model, which also had fixed effects for gender, symptom duration and reader was used as a phenotype. Three PCs derived from genotypes were used as covariates to correct for population stratification. These association results were meta-analysed using METAL (155), with weighting based on cohort size (per SNP).

Table 3.4. Accuracy of SNP2HLA imputation at selected loci. SNP2HLA was used to imputed HLA alleles for an independent cohort (the British 1958 Birth Cohort) based on a reference panel from a Type 1 Diabetes. The independent cohort and the reference panel were genotyped using the Illumina Immunochip platform. The high accuracy of imputation in an independent cohort provides validation of SNP2HLA using this genotyping platform for these loci (adapted from Jia et al. (163))

Allele	Genotyped-imputed r^2	Accuracy (concordance)
HLA-A*0201	0.99	99.4%
HLA-B27	0.97	99.9%
HLA-B40	0.96	99.7%
HLA-DRB1*0801	0.92	99.8%
HLA-DRB1*0804	0.68	99.9%

A study of Hapmap phase II data by Conrad *et al.* concluded that tag SNPs are highly portable within three broadly related ethnic groups - Africans, Asians and Europeans (166). Importantly, tag SNPs were highly portable amongst the European ethnic groups in the study. Tag SNPs from the CEU ethnic group tagged most polymorphic non-tag SNPs with $r^2 > 0.85$. Another study quantified the SNP coverage and power of CEU tag SNPs for multiethnic European populations (from Utah and Hawaii) (144). For common, non-tag SNPs, the CEU tag SNPs had high coverage (>0.87) and power (>91%).

Jia *et al.* compared reference panels from CEU and the Type 1 Diabetes Genetics Consortium with the 1958 British Birth Cohort and found imputation to be highly accurate (Table 3.4). The TASC and SPARCC cohorts were limited to subjects of European ethnicity (by questionnaire and PC analysis) in Britain, Canada, the United States and Australia. Consequently the accuracy of SNP2HLA imputation in the TASC and SPARCC cohorts should be considered similar to that of Jia *et al.*

3.2.6 Genome Wide Complex Trait Analysis

Following the lack of significant GWAS results, Genome Wide Complex Trait Analysis (GCTA) was conducted to assess whether there was any evidence that genetic variation is associated with variation in severity. Alternative methods of estimating heritability from GWA data include regression of phenotypic similarity on relatedness estimates from IBD segments (167). One advantage of this method is that it can include the effects of rare variants within these IBD segments. However IBD based methods are susceptible to error from IBD estimation in large cohorts of unrelated subjects. GCTA primarily captures the effect of common variants and may be bias towards underestimating narrow sense heritability if rare variants are significant contributors to additive genetic variation (46). GCTA estimates relatedness based on allele sharing from SNP data. One comparison of an IBD based regression method with GCTA found the standard error of its narrow sense heritability estimate to be 2-14 times as that of GCTA (46). GCTA was chosen for the heritability analysis as the purpose of this study was to assess common variants, and because its developers (who work at the same institute) were available for support.

GCTA was performed to estimate narrow sense heritability, the proportion of phenotypic variance explained by additive genetic factors. In terms of genetic factors, this method only considers additive genetic variation tagged by analysed SNPs. Consequently the estimate is expected to be lower than the true value. GCTA regresses SNP alleles against the phenotype using a linear mixed model to account for relatedness and population stratification. GCTA is able to quantify the effect of common causal variants with small effect sizes, which are usually not identified by GWAS due to stringent genome-wide significance levels. GCTA has previously been applied to height, BMI, QT interval and Parkinson's disease to demonstrate that a large narrow sense heritability is present in each case, despite not being identified by GWAS (45, 47, 168).

The GCTA used 396 921 SNPs which were genotyped in at least one of the GWAS chips. Only 78 496 of the TASC SNPs were also directly genotyped in the SPARCC cohort. For this reason imputed SNPs (see section 3.2.2) with an imputation score > 0.8 were included. The use of 400 000 SNPs for GCTA is comparable to previous studies (47,

169). Individuals with > 10% of SNPs missing were excluded, and SNPs underwent quality control as described previously (section 3.2.2). Related individuals with an Identity By State > 0.08 were excluded.

The GCTA analysis was performed only on genome-wide data from the GWAS platforms. Subjects who were genotyped using the Illumina Immunochip platform only were not included in this analysis as they had insufficient SNPs to allow GCTA – in principle and in practice. The GCTA software would not fit a model for these subjects.

The developers of GCTA recommend correcting the phenotype for known predictors and transforming it to a standardised z-score (170). The phenotype used was the ZINB model residuals, the same phenotype used in the GWAS. The phenotype of the TASC and SPARCC cohorts were standardised to z-scores then merged into one cohort. Phenotypes were regressed against the genotype using the default REstricted Maximum Likelihood (REML) method, while correcting for 3 PCs to account for subtle population structure. A per chromosome analysis was also performed.

To assess the robustness of the findings a basic sensitivity analysis used by previous authors was performed (169). The GCTA was repeated with no PCs, with 10 PCs, and with the phenotype not corrected for smoking status, symptom duration or gender.

3.2.7 Allelic Risk Score Analysis

Allelic risk score analysis was conducted to test the hypothesis that SNPs related to bone metabolism, AS susceptibility, smoking and height may also possess effects on radiographic severity in AS. Comparison of multiple loci may identify variation from a collection of markers which in isolation do not achieve significance (171). This method is useful for testing candidate SNPs when no individual locus is found to be significant upon independent association testing. In this study, combining the effects of many SNPs into a single score increase the power to detect a small effect in an study otherwise poorly powered to detect independent effects across millions of markers. A positive result would provide some signal that, for example, variation in bone metabolism loci can predispose more severe disease. Such a signal would justify a candidate gene study of these loci.

Allelic risk scores are limited by a number of assumptions, including that: the score contains SNPs in LD with causal variants, that many (and not few) alleles included in the score are in LD with causal variants and that the effect size is similar between the two phenotypes. These limitations are discussed at length in section 3.4.4.

Allelic risk scores were calculated to investigate whether SNPs associated with AS susceptibility, bone mineral density (BMD), height and nicotine dependence also predispose to AS severity. Creation of an allelic risk score using SNPs associated with a related phenotype has been described previously (172). For example polygenic risk scores have been used to predict blood pressure from an allelic risk score derived from the related phenotype of Body Mass Index (173).

SNPs associated with AS susceptibility were identified from previous studies of AS susceptibility (22, 35, 36). SNPs associated with BMD, height and nicotine dependence were identified by GWAS of these traits (174-177).

In the case of nicotine dependence, two significant GWAS were identified, published in 2007 and 2010 respectively (175, 177). Without a meta-analysis, it would be difficult to

combine hits from each GWAS into a single score with a common effect size. The latter, larger GWAS associated three SNPs with nicotine dependence. Only two of these SNPs were imputed in the TASC or SPARCC cohorts, each of which tags a nicotinic acetylcholine receptor subunit locus. One of these tag SNPs is in perfect LD with rs13277254, a SNP identified by the 2007 GWAS. Thus for the purpose of choosing alleles for this allelic risk score analysis, the additional risk allele from the latter study adds little to the earlier 2007 study. The latter GWAS only contributes one unique allele to the risk score, which corresponds to a subunit of a receptor that is tagged by other SNPs in the 2007 study. For this reason, only risk alleles from the 2007 nicotine dependence GWAS were used.

One SNP was included per risk locus – the SNP with the highest effect size. For each subject, the number of risk alleles present was counted, and multiplied by the estimated effect size of the risk allele and aggregated to provide a weighted allelic risk score. The effect size listed by the AS susceptibility GWAS and nicotine dependence GWAS were Odds Ratios (ORs) as they were case-control studies. The BMD GWAS and height GWAS listed effect size as the linear regression coefficient β, as they were quantitative GWAS. For the allelic risk score calculation, Odds Ratio (OR) effect sizes were converted to β coefficients by logarithmic transformation. Only high confidence imputed SNPs with imputation score >0.8 were used. Subjects missing > 5% of the SNPs used in that score were excluded from that analysis. Separate allelic risk scores were created for AS susceptibility, BMD, height and nicotine dependence.

For the BMD cohort, the effect size used was the mean effect size for femoral and lumbar spine BMD for each risk allele.

HLA-B27 was not used in the analysis using the AS susceptibility allelic risk score, as its OR of 60 made all other SNPs irrelevant to the score. The SNP with the next highest effect size was rs11209026 within *IL23R* with OR 1.6.

Each score was then assessed for predictive ability by using it as a predictor in the ZINB model. SPARCC subjects were analysed separately to TASC subjects as one cohort had

mSASSS data and the other cmSASSS. The ZINB models (apart from the nicotine dependence score model) were corrected for the influence of symptom duration, gender and three level smoking status. The nicotine dependence score model was not corrected for smoking status in any way, allowing the nicotine dependence score to influence severity via smoking. In the case of nicotine dependence, only males were included in the allelic risk score. The nicotine dependence paper listed separate OR for males and females, and there were too few females with risk scores for the model to calculate a result for them.

Table 3.5. Number of SNPs used for allelic risk scores. Some risk loci could not be included because their risk SNPs were not imputed or genotyped.

Allelic Risk Score	SNPs in TASC cohort	SNPs in SPARCC cohort
AS susceptibility	41	41
BMD	83	81
Height	133	126
Nicotine dependence	40	40

3.3 Results

3.3.1 *A priori* Power Calculation Results

Power was estimated over a range of trait heritabilities. For interpretation, heritability was converted to a range of allele frequencies and effect sizes.

The effect size cannot be converted to meaningful uncorrected (ie observed) cmSASSS or syndesmophytes, but represents variation in the phenotype after correcting for smoking status, gender and symptom duration. The effect size should be considered as an additive effect present in each subject. For example, an individual with 3 SNPs with effect size +0.1 would be expected to have a cmSASSS 0.3 standard deviations above the mean, after correcting for the effect of symptom duration, gender and smoking.

The GWAS was powered to identify common SNPs with a large effect size. The GWAS had 80% power to detect SNPs with MAF of 0.2 with an effect size of 0.49 standard deviations from the mean (of cmSASSS ZINB residuals) to the level of significance $\alpha = 5 \times 10^{-8}$. Standardised effect sizes identified by similar GWAS have been much smaller than 0.49. The GEFOS2 study of BMD found dozens of SNPs with standardised effect sizes <0.1 and a few SNPs with standardised effect sizes of 0.1 – 0.15 (176). The GIANT study identified 180 loci associated with height ($p < 5 \times 10^{-8}$), yet only one locus had standardised effect size > 0.1. The interpretation of standardised effect sizes suggested by Cohen for general statistics, where 0.2 is small, 0.5 moderate and 0.8 large, is therefore not applicable to quantitative GWAS, as the effect sizes that are actually observed are much smaller than these thresholds. Thus the GWAS was only powered to identify common SNPs with large effect sizes, in regards the known likely effect sizes of common genetic variants on bone phenotypes.

Figure 3.7. Power for 1360 subjects. D'=0.8, α= 5 x 10^{-8}. The dotted line represents a power of 0.8 (80%). For common variants, only large effect sizes and common MAF (>0.05) are powered > 0.8.

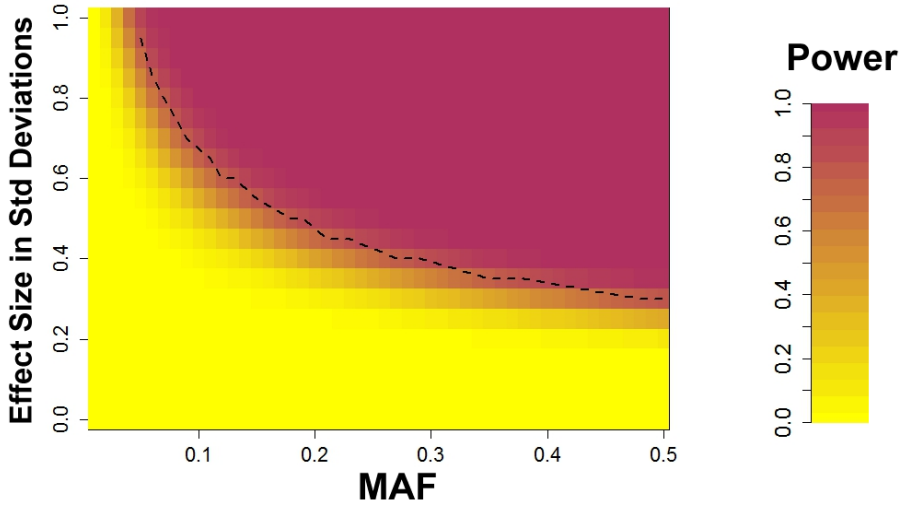


Figure 3.8. Power for 3000 subjects. D'=0.8, α= 5 x 10^{-8}. Doubling the sample size slightly improves the likelihood of detecting alleles with lower effect sizes.

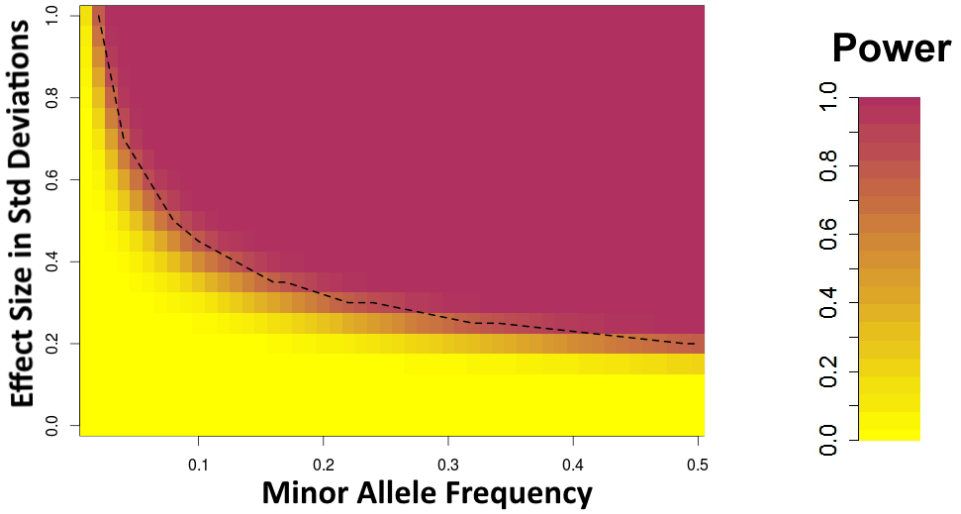
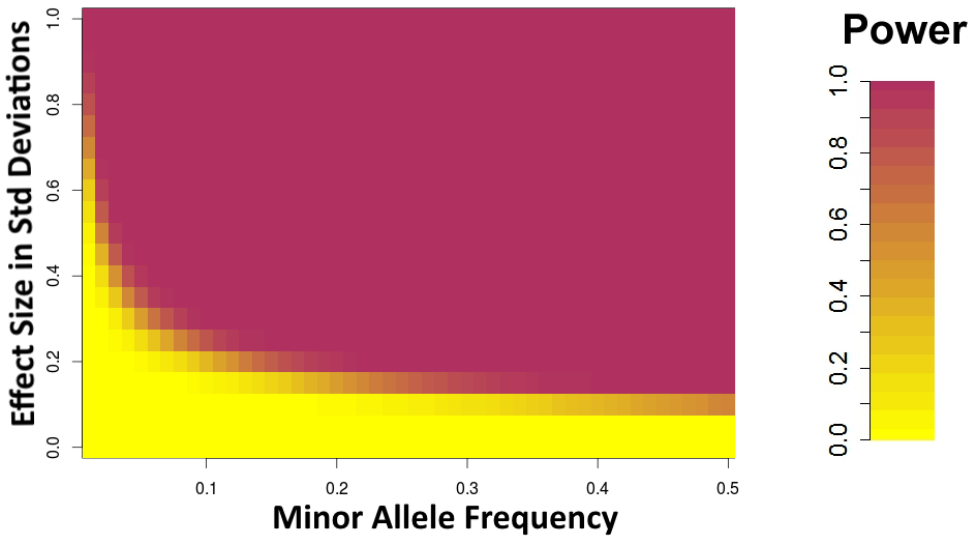


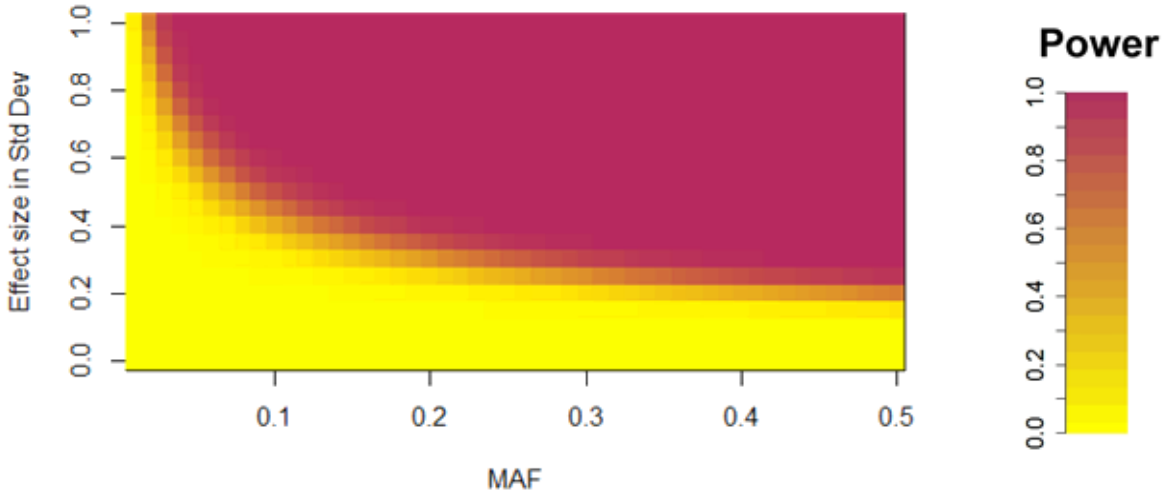
Figure 3.9. Power for 10 000 subjects. D'=0.8, α= 5 x 10^{-8}. This large sample size still provides low power for effect sizes < 0.1 standard deviations.



Power calculations demonstrate that even with a sample size of 10 000, there would only be adequate power for small effect sizes (<0.1) if MAF > 0.4.

Power calculations for the HLA imputation step were performed 7674 SNPs from 1144 and 876 subjects. Similarly association testing would only be well powered to detect large effects in common variants.

Figure 3.10. Power for 1144 subjects. D'=1, α = 6.5 x 10^{-6}. α is based on a Bonferroni correction for 7674 independent tests, assuming no LD. The assumption of no LD for this HLA testing makes the correction overly stringent.

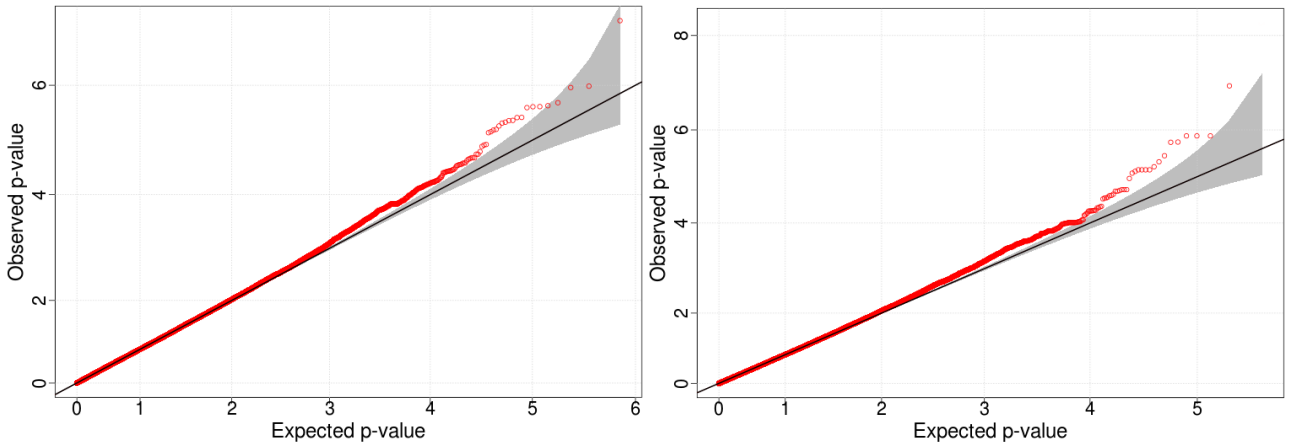


3.3.2 Genome Wide Association Study Results

Following quality control and exclusion of subjects without mSASSS, symptom duration, gender, or smoking status, 979 subjects from the TASC cohort and 274 subjects from the SPARCC cohort were included. For the TASC cohort, Genomic Inflation Factor (λ) = 1.01 and for the SPARCC cohort, λ = 1.00. Neither cohort had a SNP more significant than $p < 1 \times 10^{-7}$. SNP effect size in each cohort was approximately normally distributed ($\mu = 0$, $\sigma = 1.7$) with more extreme values than the normal distribution in the top and bottom quartiles. Standardised effect sizes from the TASC cohort are reported for comparison.

After removal of rare SNPs (MAF < 0.05), 5 657 478 SNPs remained. These SNPs were analysed by p-value meta-analysis using METAL (155).

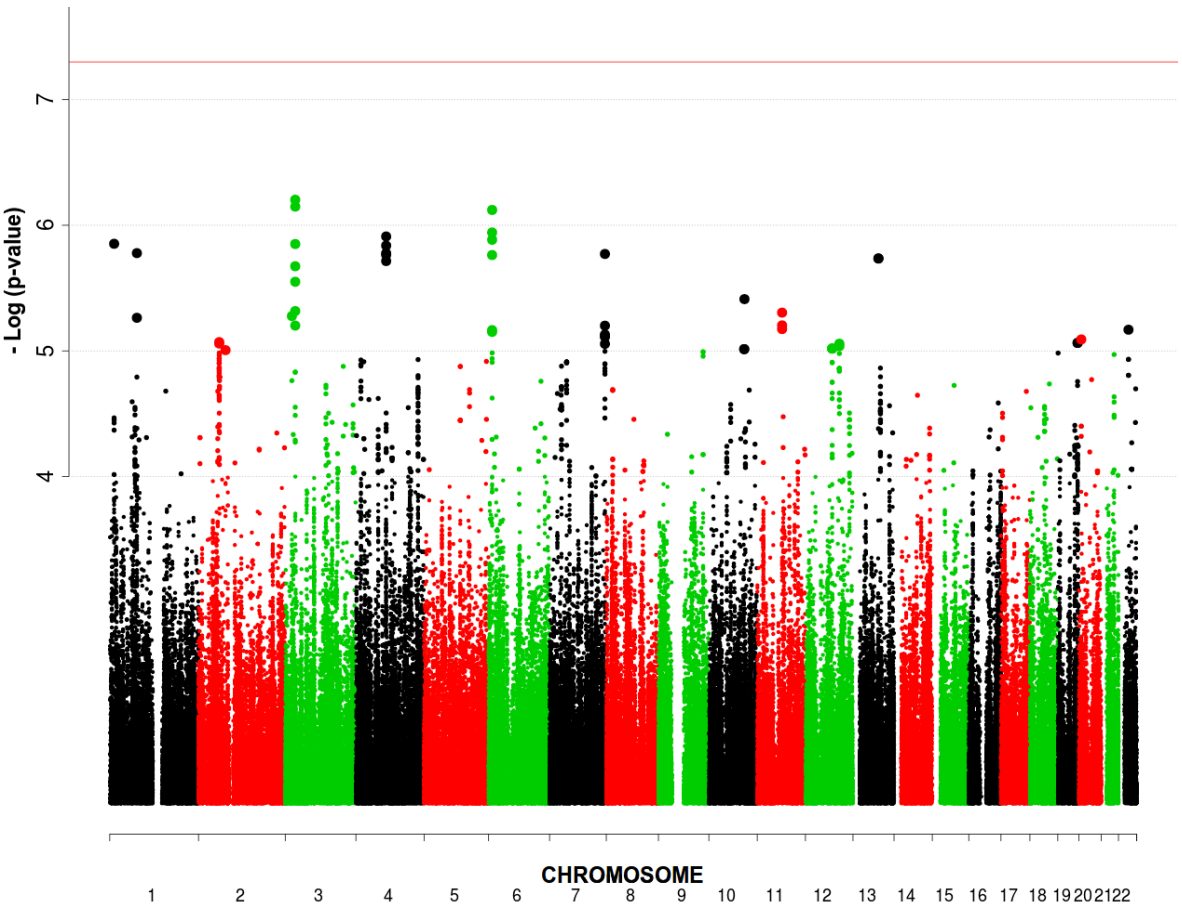
Figure 3.11. QQ plots for each cohort. SPARCC GWAS (left) and TASC GWAS (right). There is minimal inflation compared to the expected values by chance (black line, with 95% confidence interval in grey). Very few observed p-values were < 1×10^{-5}.



After meta-analysis, 92.3% of SNPs were present in at least 800 subjects, while 507 910 SNPs were present in more than 1000 subjects. No SNP with MAF ≥0.05 met the

experiment wide significance cutoff $p < 8.8 \times 10^{-9}$, based on a Bonferroni correction for 5.66 Million tests, or the arbitrary genome wide significance cutoff of 5×10^{-8} (Bonferroni for 1 million SNPs).

Figure 3.12. Meta-analysis Manhattan plot. The red line is at $p = 5 \times 10^{-8}$. The large centromere of Chromosome 9 is visible.



Significant associations within the HLA region were not examined using this genome wide approach because the SNP2HLA program imputed the HLA region with much higher accuracy.

Suggestive associations ($p < 1 \times 10^{-6}$) were found at 10 loci. A suggestive association was present at a TNFα receptor gene, Tumour Necrosis Factor Receptor Superfamily, Member

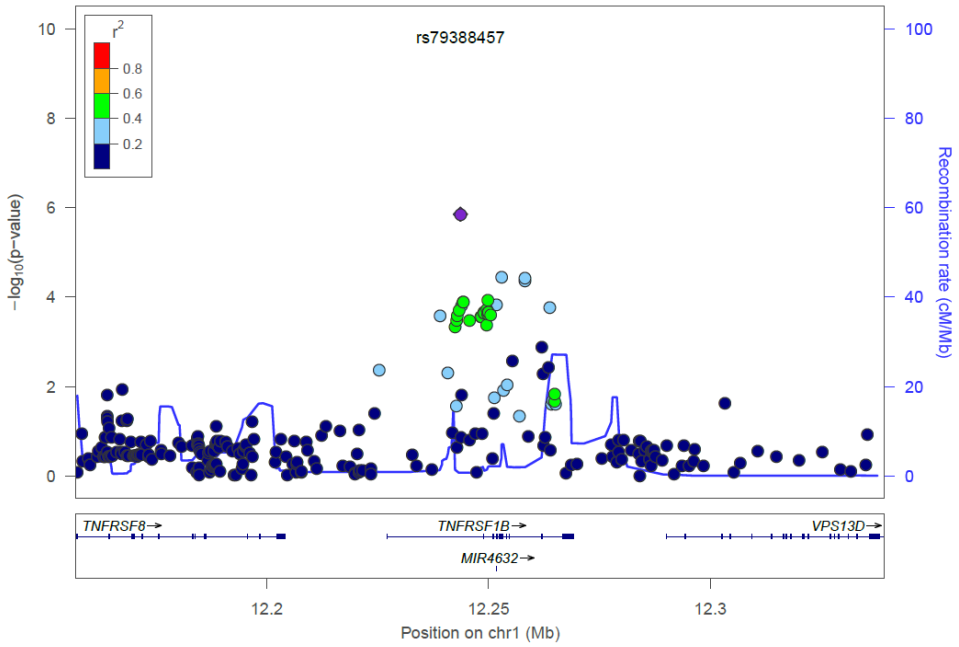
1B (*TNFRSF1B*). The strongest association was at SNP rs79388457 which is located in an intron. This gene is expressed in leucocytes. Unlike *TNFRSF1A*, it has not been identified as a significant association in previous AS susceptibility GWAS or previous candidate gene studies of radiographic severity.

Table 3.6. TNFRSF1B association results. Association was present in imputed SNPs from both cohorts and from a genotyped SNP. The standardised effect size of the TASC SNP rs79388457 is 3.56.

SNP	Position (base pairs)	MAF	Subjects with SNP value and SNP type	Effect Direction	SPARCC cohort p-value	TASC cohort p-value	Meta-analysis p-value
rs79388457	12243674	0.13	877 Imputed	++	0.09	3.9×10^{-6}	1.4×10^{-6}
rs235221	12263817	0.32	211 Imputed	++	1.7×10^{-4}	0.03	1.7×10^{-4}
rs1061622	12252955	0.23	877 Genotyped	++	2.7×10^{-3}	2.2×10^{-3}	3.5×10^{-5}

*Risk Allele Frequency from 1000 Genomes dataset

Figure 3.13. Regional association plot for TNFRSF1B (meta-analysed SNPs). The strongest association signal is from the intronic SNP rs79388457.



An association signal was present at the NOD-Like Receptor Pyrin domain-containing protein 7 *(NLRP7* also known as *NALP7, NOD12, PYPAF3)* locus, with peak association 5 kb from the gene. It is expressed in leucocytes and gastrointestinal and respiratory epithelium and plays a role in inflammasome regulation. The lead SNP is between *NLRP7* and Natural Cytotoxicity Triggering Receptor 1 (*NCR1*), a receptor capable of triggering NK cell activation. The lead SNP is in an intergenic region but many of the significantly associated imputed SNPs (from both cohorts) are within *NLRP7*, making *NLRP7* the more likely source of association.

Table 3.7. NLRP7 association results. Association was present in both cohorts and in a SNP genotyped in >1000 subjects. Notably the lead meta-analysed SNP is a low frequency variant with MAF <0.1, and has a large effect size in the TASC cohort. The standardised effect size of the TASC SNP rs139325508 was 4.01.

SNP	Position (base pairs)	MAF	Subjects with SNP value and SNP type	Effect Direction	SPARCC p-value	TASC p-value	Meta-analysis p-value
rs139325508	55429591	0.06	974 Imputed	++	0.31	6.3×10^{-6}	8.6×10^{-6}
rs11668142	55900417	0.39	226 Imputed	--	4.8×10^{-4}	0.15	2.6×10^{-3}
rs269933	55439166	0.46	1035 Genotyped	--	0.2	1.3×10^{-4}	6.6×10^{-5}

Figure 3.14. Regional association plot for NLRP7 (meta-analysed SNPs).

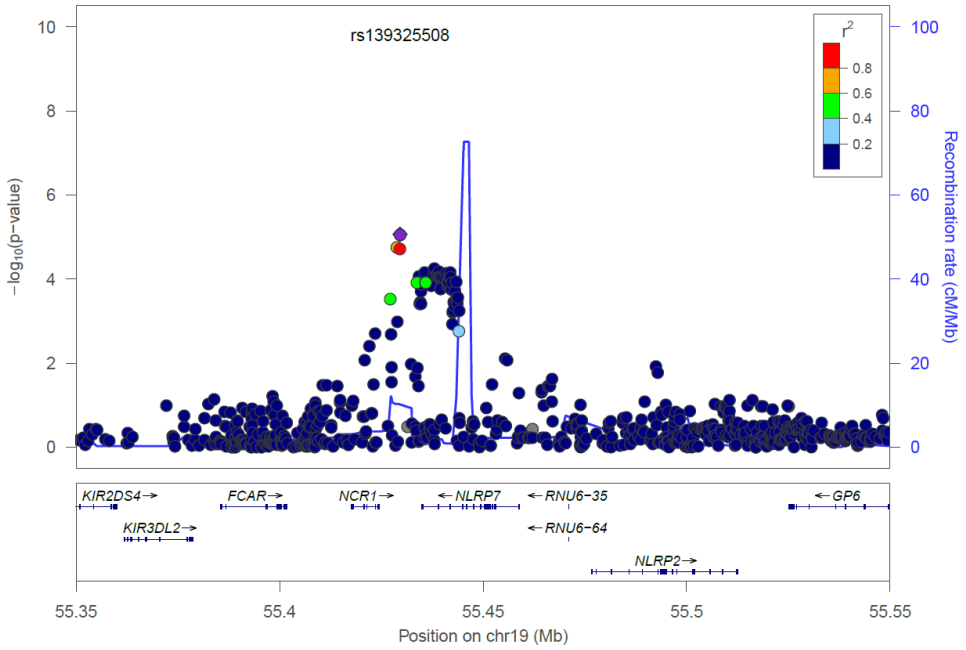
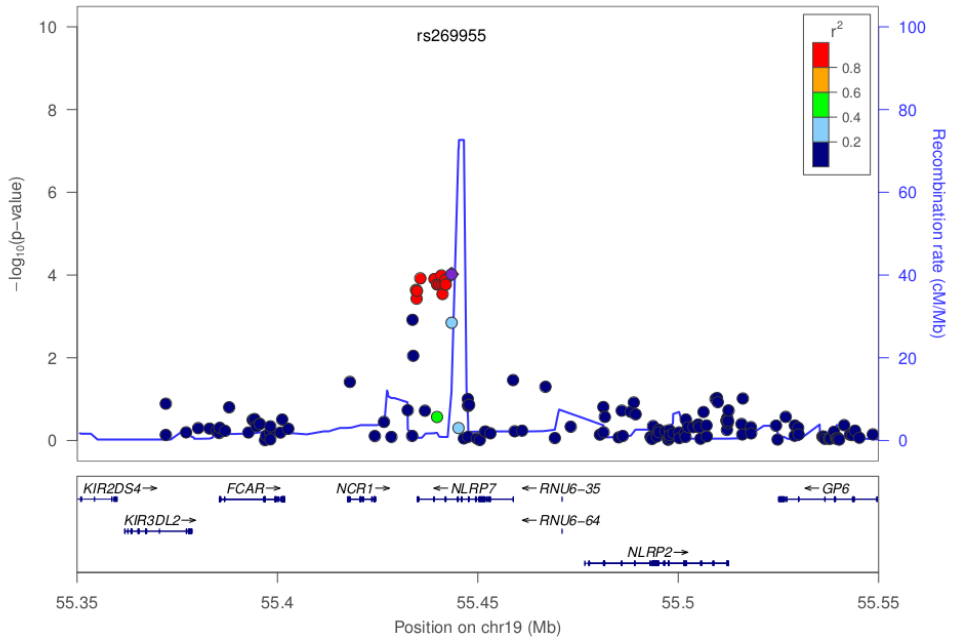


Figure 3.15. Regional association plot for NLRP7. Genotyped SNPs from TASC cohort. Multiple independently genotyped SNPs are in the association peak ($p \sim 10^{-5}$). Red indicates that these SNPs are in high LD with the lead SNP (purple).

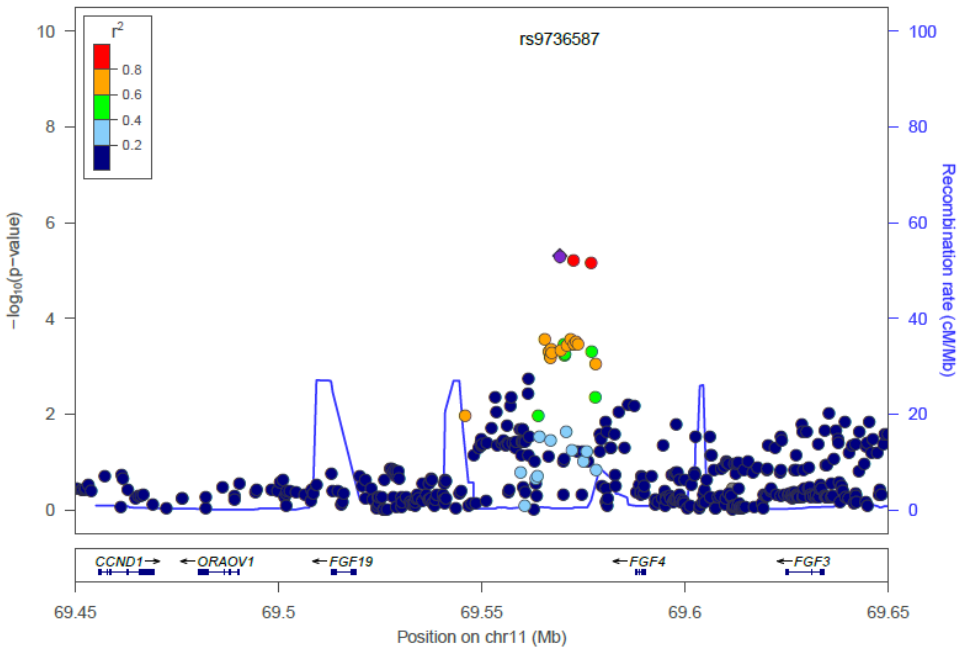


A suggestive association was seen near the gene Fibroblast Growth Factor 4 (*FGF4*), with peak association 21kb from the final exon. The peak could represent an insulator or enhancer of *FGF4*. Alternately the locus could regulate *FGF19*, which is 50kb from the association peak. *FGF4* is expressed in skeletal muscle, tendon, granulocytes, osteoblasts and respiratory epithelium.

Table 3.8. FGF4 association results. A suggestive association was present upon meta-analysis, but the most significant SNPs in each cohort and from genotyped data had p-values of magnitude 10^{-4} and 10^{-3}. The standardised effect size of the lead TASC SNP rs1944131 was 1.84.

SNP	Position (base pairs)	MAF	Subjects with SNP value and SNP type	Effect Direction	SPARCC cohort p-value	TASC cohort p-value	Meta-analysis p-value
rs1944131	69572577	0.32	603 Imputed	++	0.01	1.6×10^{-4}	Not avail
rs73518063	69879369	0.06	211 Imputed	--	1.2×10^{-3}	Not avail	Not avail
rs9736587	69569224	0.31	Geno: 603 Imp: 485	++	0.01		5.0×10^{-6}

Figure 3.16. Regional association plot of FGF4 (meta-analysed SNPs). The association peak demonstrates LD with *FGF4*, demonstrated by the recombination peak (peaked blue line) at 69.6 Mb.



A suggestive association was seen within Bone Morphogenic Protein 2 (*BMP2*). This association was only of magnitude 10-3 in each cohort but is notable because of the known involvement of BMP2 in bone metabolism. Peak association was within an exon. All meta-analysed SNPs with p <1 x 10^{-4} had MAF of 0.07 in this cohort. For example, the lead meta-analysed SNP had a Risk Allele Frequency 0.07, so would be present in 57 individuals in this study. The effect size of the lead TASC SNP is high, and of similar magnitude to the lead TASC *TNFRSF1B* SNP. The small sample size of this GWAS, and peak association in genotyped SNPs of only 5 x 10^{-3} makes this association tentative.

This association was only of magnitude 10^{-3} in each cohort but is notable because of the known involvement of BMP2 in bone metabolism.

Table 3.9. BMP2 association results. Association at BMP2 was present in both cohorts, with the strongest association in low frequency variants with MAF <0.1. The standardised effect size of the lead TASC SNP rs13037675 was 5.9.

SNP	Position (base pairs)	MAF	Subjects with SNP value and SNP type	Effect Direction	SPARCC cohort p-value	TASC cohort p-value	Meta-analysis p-value
rs13037675	6759706	0.07	814 Imputed	++	0.037	7.8×10^{-5}	8.1×10^{-6}
rs11697474	6722811	0.07	211 Imputed	++	5.4×10^{-3}	Not avail	Not avail
rs6910011	6877411	0.45	211 Genotyped			5.1×10^{-3}	5×10^{-3}

Figure 3.17. Regional association plot of BMP2 – meta-analysed SNPs.

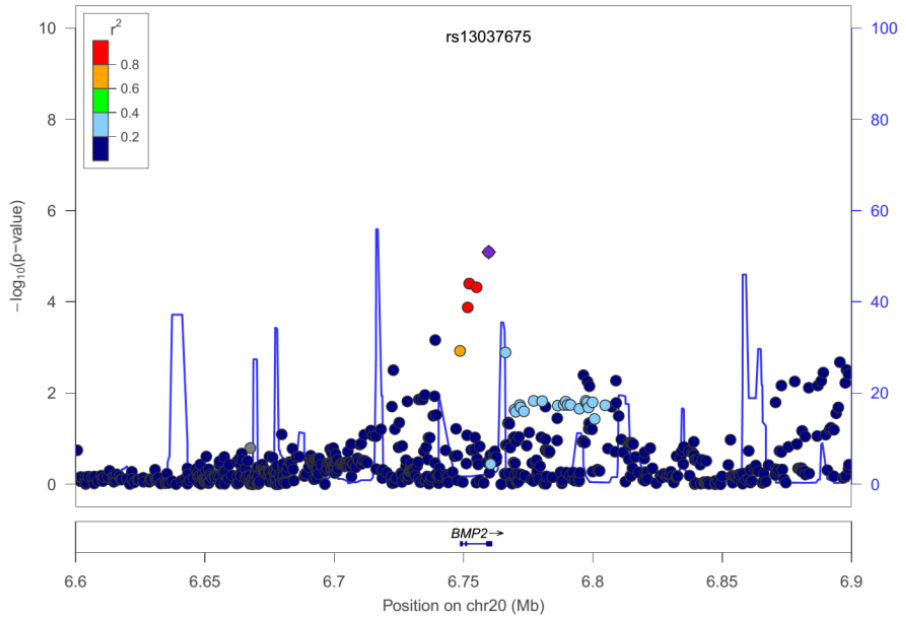
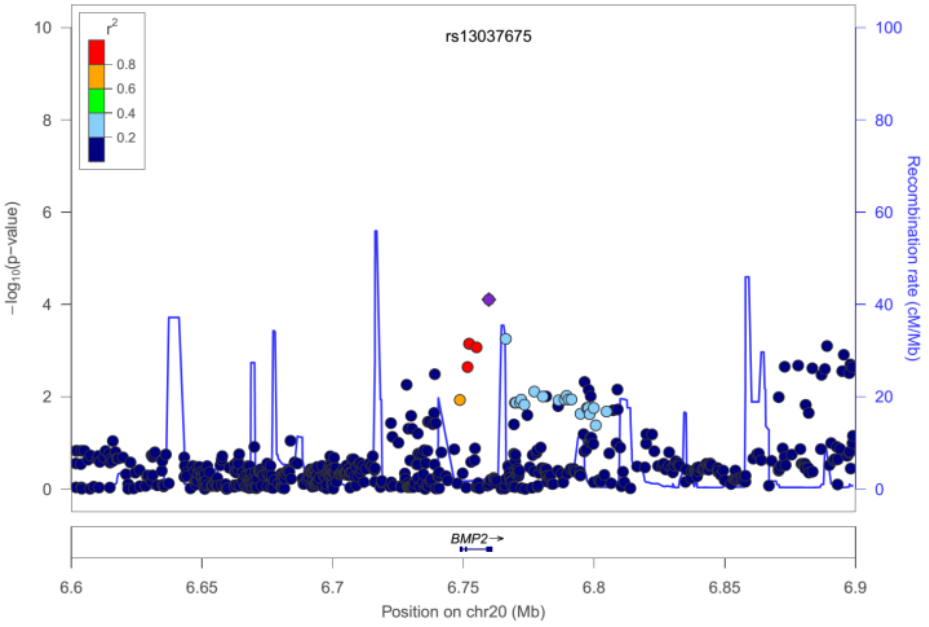
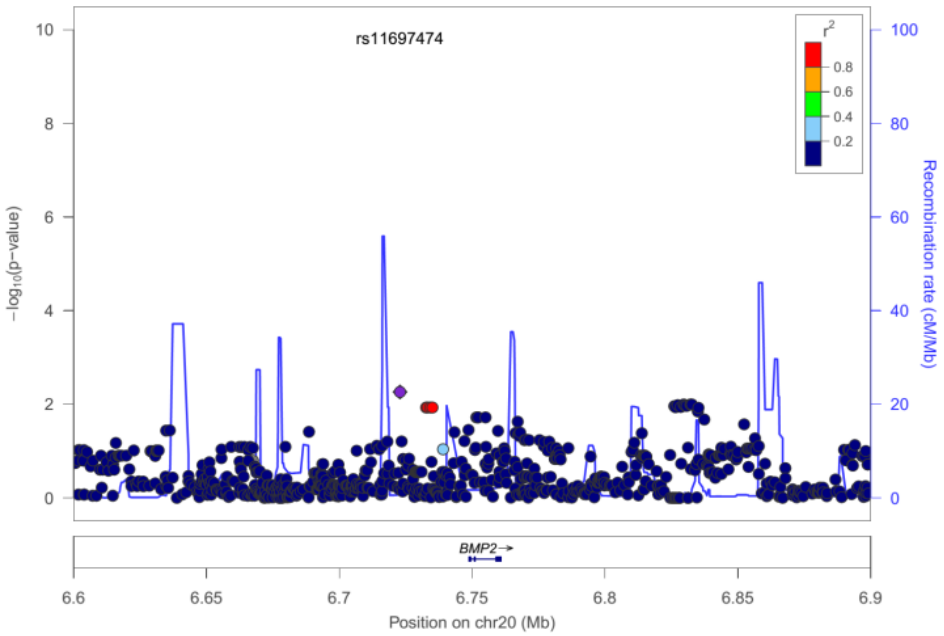


Figure 3.18. BMP2 regional association plot by cohort. TASC imputed SNPs (A) and SPARCC imputed SNPs (B). Little association is seen in the SPARCC cohort.

A)



B)



A suggestive association was seen at Orofacial Cleft Candidate 1 *(OFCC1)*. Chromosomal breakpoints at this locus have repeatedly been associated with congenital orofacial cleft in humans, however *OFCC1* knockout does not cause this malformation in

mice (178). *OFCC1* is expressed by respiratory epithelium, stimulated monocytes, bone marrow stromal cell lines and smooth muscle. The peak association at OFCC1 is located in an intron.

Table 3.10. OFCC1 association results. Association at *OFCC1* is present in both cohorts. The standardised effect size of lead TASC SNP rs9383171 is 4.19.

SNP	Position (base pairs)	MAF	Subjects with SNP value and SNP type	Effect Direction	SPARCC cohort p-value	TASC cohort p-value	Meta-analysis p-value
rs9383171	9873422	0.17	1088 Imputed	--	0.18	4.4×10^{-7}	1.1×10^{-6}
rs6914253	9824464	0.14	814 Imputed	++	8.2×10^{-6}	0.65	7.7×10^{-3}
rs9396671	9869358	0.12	1088 Genotyped	--	0.11		7.6×10^{-7}

Figure 3.19. Regional association plot of OFCC1 (meta-analysed data).

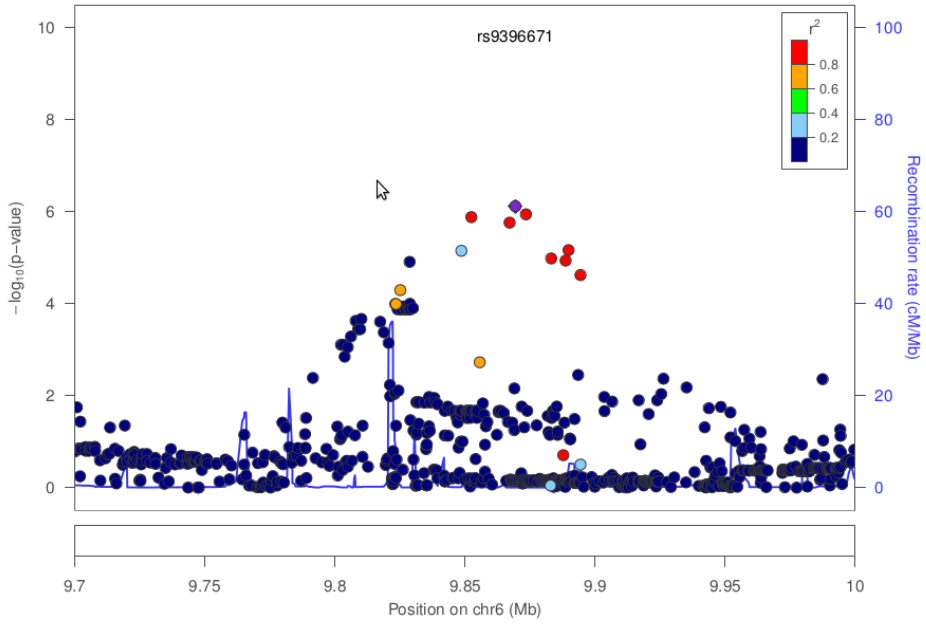


Figure 3.20. Regional association plot of OFCC1 (TASC cohort). The association peak extends to either side of a recombination hotspot (peaked blue line). Consequently there are two non-linked signals within *OFCC1* (coloured red and navy), one on each side of the recombination hotspot.

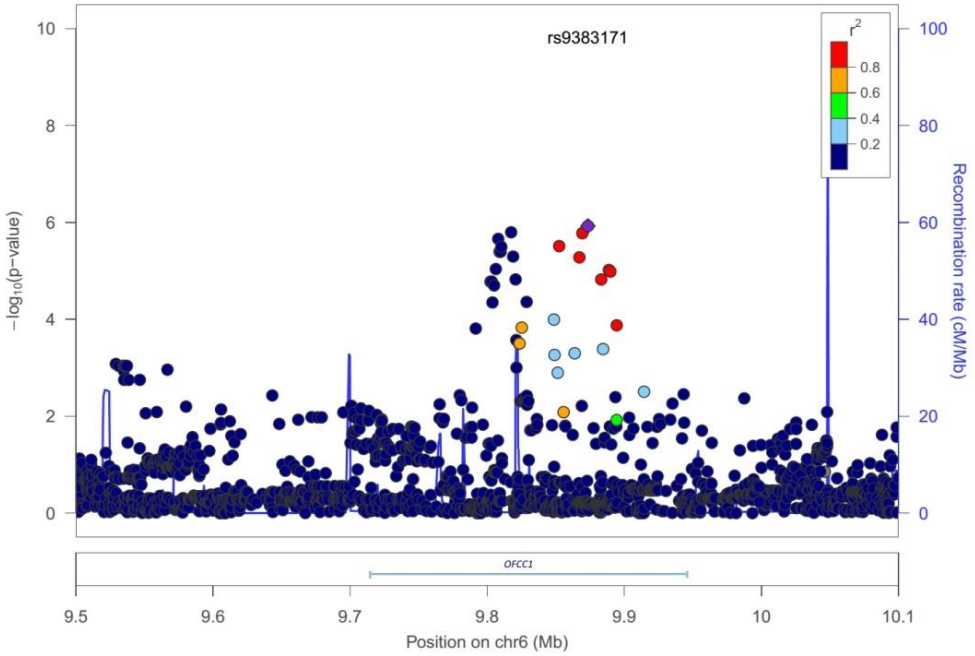
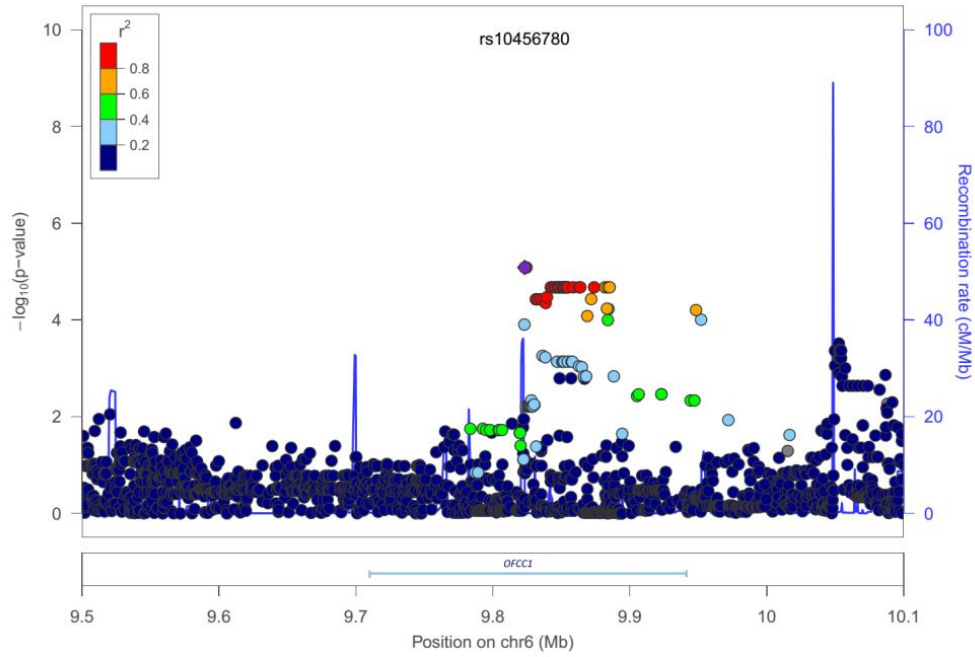


Figure 3.21. Regional association plot of OFCC1 (SPARCC cohort). The association peak is independently present the SPARCC cohort too.



A suggestive association is present at the protein kinase Vaccinia Related Kinase 2 (*VRK2*). Peak association is present in an intron. *VRK2* modulates transcription of cytokines including IL-1β and has been associated with schizophrenia and generalised epilepsy (179, 180). *VRK2* is expressed in stimulated monocytes, neutrophils, tendon and the substantia nigra. Two independently genotyped SNPs in the TASC cohort were significant p ~ 10^{-6}.

Table 3.11. VRK2 association results. Association was predominantly in the TASC cohort, in imputed and genotyped SNPs. The SPARCC imputed SNP is a low frequency variant with MAF <0.1. The standardised effect size of lead TASC SNP rs2716999 is 2.36.

SNP	Position (base pairs)	MAF	Subjects with SNP value and SNP type	Effect Direction	SPARCC cohort p-value	TASC cohort p-value	Meta-analysis p-value
rs2716999	58075743	0.3	603 Imputed	++	0.48	1.1×10^{-6}	8.7×10^{-6}
rs144500729	57003489	0.06	211 Imputed	--	2.7×10^{-3}	Not avail	2.6×10^{-3}
rs1607889	58073603	0.3	603 Genotyped	++	0.55		1.8×10^{-6}

Figure 3.22. VRK2 regional association plot. (meta-analysed data). The strong p-values in the TASC cohort were diminished by the lack of association in the SPARCC cohort. The closest gene in the 5' direction to the association peak is *CCDC85A*, an uncharacterised gene 1.5 MB distant.

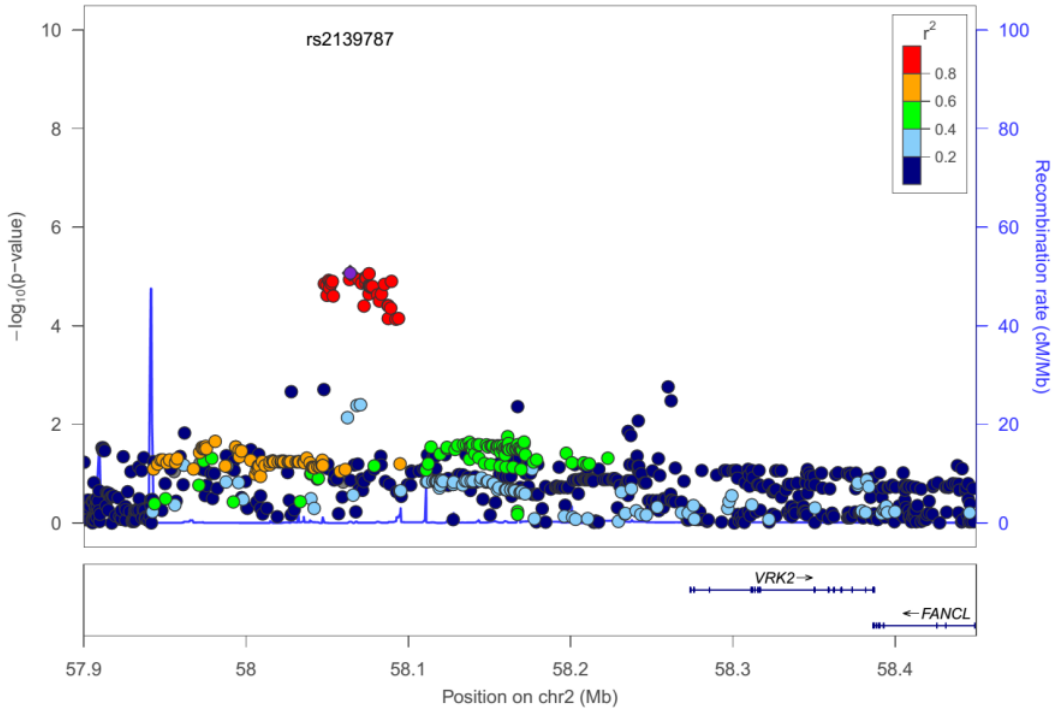
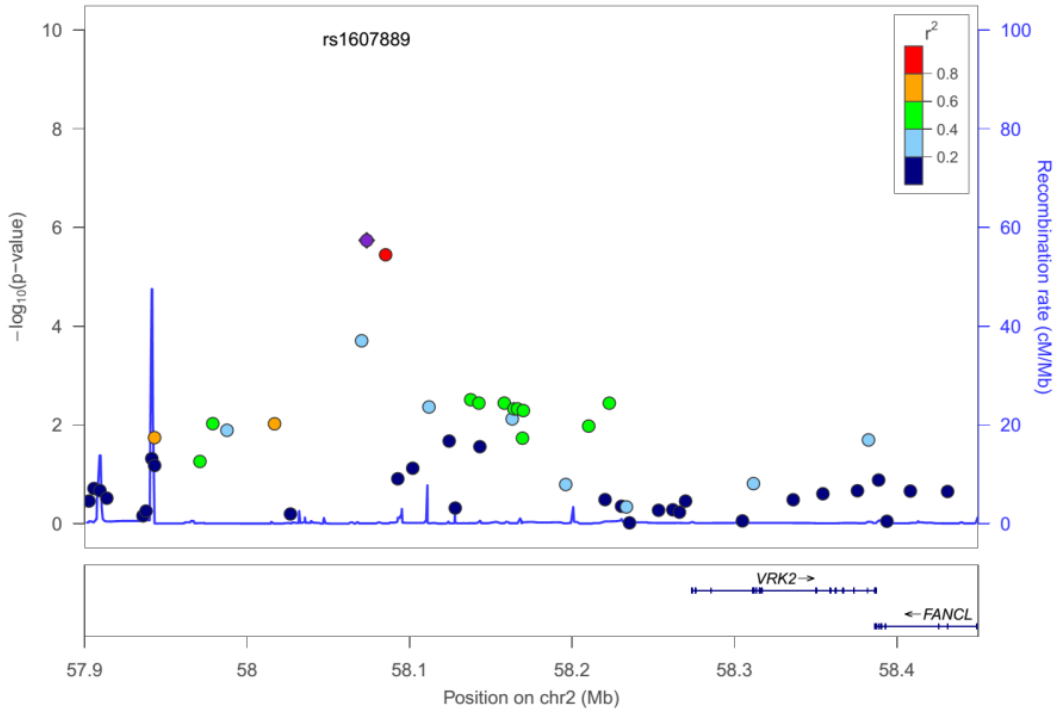


Figure 3.23. VRK2 regional association plot (*genotyped TASC SNPs*). This plot consists of genotyped SNPs from the TASC cohort only, and appears sparse compared to the imputed regional association plots. Notably, two linked SNPs (purple and red) had p-values of magnitude 10^{-6}.



A suggestive association was seen at 3p24.1. Peak association was 216 bp upstream of a genscan predicted transcript and 23.5 kb downstream an ENSEMBL predicted gene (ENSG00000235493). The transcripts at this locus are predicted to code for long intergenic non-coding RNA (lincRNA). This peak is 255 kb from the gene eomesodermin (*EOMES*), which is a known AS susceptibility locus (22). No regulatory features such as DNA methylation were present at this locus. The association was strongest in low frequency and rare variants.

Table 3.12. 3p24.1 association results. Association at this locus was predominantly in TASC subjects, in low frequency variants with MAF <0.1. Genotyped TASC SNPs and SPARCC subjects only had association of 1 x 10^{-3}. The standardised effect size of lead TASC SNP rs77367404 is 3.18.

SNP	Position (base pairs)	MAF	Subjects with SNP value and SNP type	Effect Direction	SPARCC cohort p-value	TASC cohort p-value	Meta-analysis p-value
rs77367404	28015174	0.08	603 Imputed	++	0.01	1.6 x 10^{-5}	6.3 x 10^{-7}
rs7623233	27922474	0.12	211 Genotyped	++	1.0 x 10^{-3}	0.31	0.01

A suggestive association was seen at 4q21.23, in the intron of ENSEBML predicted gene ENSG00000250546. The transcript is expressed in muscle, leucocytes, fibroblasts and chrondrocytes, and is predicted to be lincRNA. This pattern of expression is consistent with sites of inflammation and bone formation in AS.

Table 3.13. 4q21.23 association results. Association was present in each cohort and in a genotyped SNP. The standardised effect size of the lead TASC SNP rs11099621 is 1.73.

SNP	Position (base pairs)	MAF	Subjects with SNP value and SNP type	Effect Direction	SPARCC cohort p-value	TASC cohort p-value	Meta-analysis p-value
rs11099621	84762151	0.4	Geno: 603 Imp: 211	--	0.002	1.4 x 10^{-4}	1.23 x 10^{-6}
rs13147593	84756536	0.16	211 Imputed	++	6.8 x 10^{-4}	0.02	2.3 x 10^{-4}

A suggestive association was seen 82.5 kb from the last exon of 5-Hydroxytryptamine receptor 5A (*HTR5A*), a G-protein coupled receptor. This gene is expressed in multiple cell types of the central nervous system.

Table 3.14. HTR5A association results. Association was present in both cohort and in genotyped SNPs. The standardised effect size of the lead TASC SNP rs10260348 is 2.26.

SNP	Position (base pairs)	MAF	Subjects with SNP value and SNP type	Effect Direction	SPARCC cohort p-value	TASC cohort p-value	Meta-analysis p-value
rs10260348	154959965	0.45	814 Imputed	--	0.19	1.7×10^{-6}	1.7×10^{-6}
rs1973359	154135286	0.18	211 Imputed		1.4×10^{-3}	0.93	0.09
rs1730175	154955663	0.29	814 Genotyped	--	0.017	2.4×10^{-4}	1.3×10^{-5}

3.3.3 HLA Association

7674 HLA SNPs were imputed for 1144 subjects who were genotyped on Immunochip. Association testing was performed using three level smoking status model residuals for each cohort, then meta-analysed. Only 876 subjects had a three level smoking status.

No HLA haplotype had a signification or suggestive association. *HLA-B27* had no significant association (p=0.34). No HLA alleles or amino acid polymorphisms had a suggestive association, and no allelic association met significance using the Bonferroni correction for 7674 independent tests of 6.5×10^{-6}. The most significant p-value was 3.7×10^{-4} for the intergenic SNP rs12203825. No HLA allele listed by a previous candidate gene study by Ward *et al.* was significant with $p < 0.1$ (71).

The process was repeated ignoring correction for smoking status, allowing 1144 subjects to be included. Again, no significant association was found. Peak association was at the intergenic SNP rs3130320 $p= 2.0 \times 10^{-4}$.

Figure 3.24. Regional association plot at HLA B and C loci. Only SNP associations are plotted – amino acid and HLA association results are not shown. Not significant SNP association is present. The nearest suggestive association is at OFCC1 (not shown).

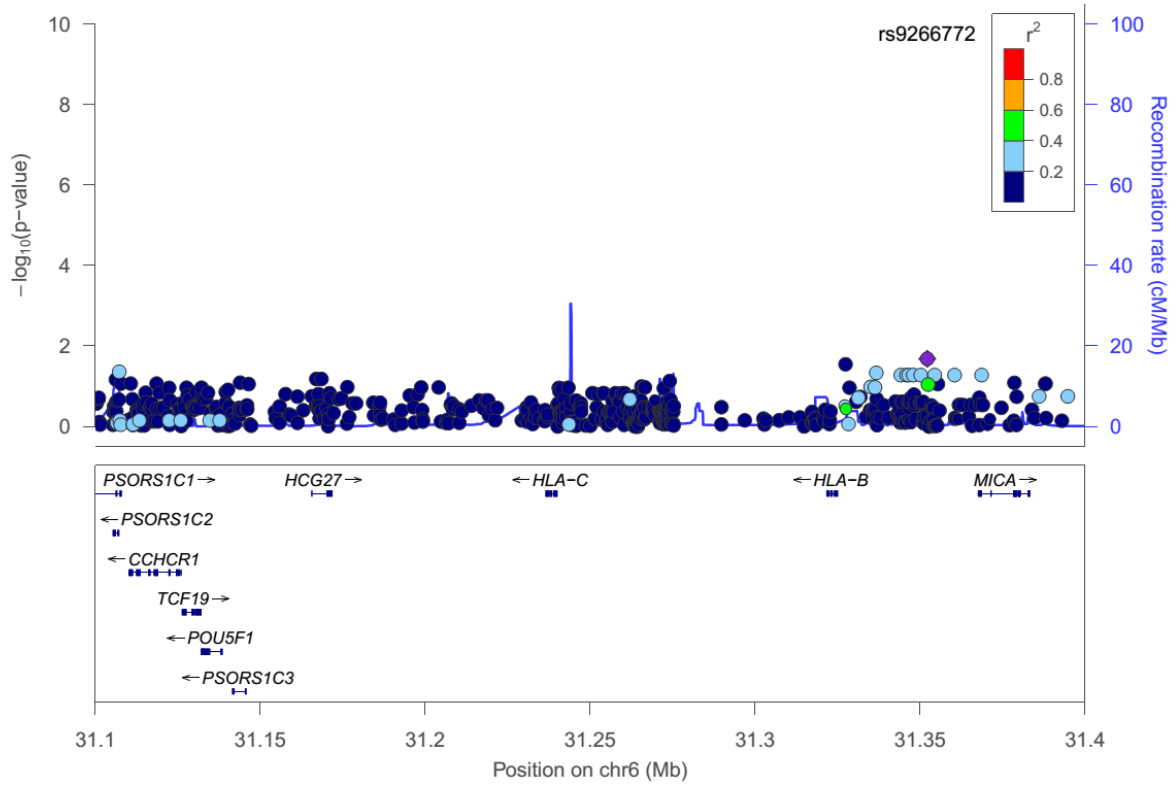
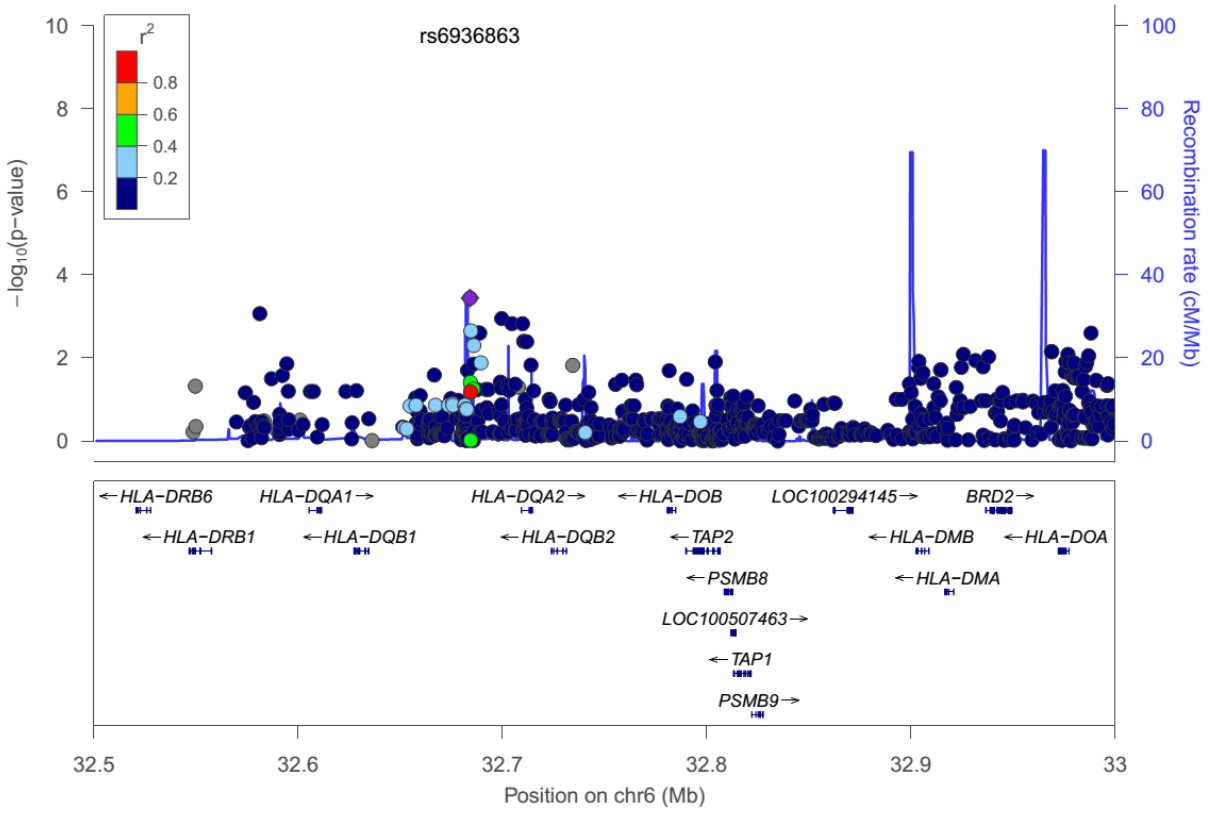


Figure 3.25. Regional association plot at HLA Class II loci. No significant SNP association is present.



3.3.4 GCTA Results

For the phenotype of mSASSS and cmSASSS, corrected for gender, symptom duration and smoking status, the estimated phenotypic variation due to genetic factors (V_G / V_P) was 0.330 or 33% (p = 0.006, standard error 0.136, using 984 subjects).

The per chromosome analysis was limited by the small number of subjects and SNPs. Only two chromosomes had significant results – Chromosomes 3 and 5. Chromosome 6, which includes the HLA, did not have a significant V_G / V_P.

Table 3.15. Significant per chromosome GCTA results.

Chromosome	SNPs included	Significance	V_G/V_P	Standard Error
3	20 414	p = 0.02	0.082	0.05
5	18 070	p = 0.01	0.102	0.054
18	9 857	p = 0.05	0.098	0.055
6	31 497	p = 0.20	0.023	0.034

The basic GCTA sensitivity analysis demonstrated that the results were robust to confounding covariates including population structure and model covariates. The phenotype was found to be significantly heritable without correction for gender or smoking status (p = 0.03), without correction for PCs (p = 0.01) or while correcting for ten PCs (p = 0.01).

3.3.5 Allelic Risk Score

In the both cohorts, allelic risk scores for AS susceptibility, nicotine dependence and height were not predictive of radiographic severity. In the SPARCC cohort only, the BMD score was a significant predictor of mSASSS, with a borderline level of significance (p=0.037). The SPARCC model reported a large positive effect size of 0.67 standardised units for BMD score, slightly greater than that of gender at 0.58 units. In the TASC cohort the BMD score was not a significant predictor and had a positive effect size of 0.25 units. If BMD score were a true predictor of magnitude greater than gender, it should be significant in both cohorts. For this test, the TASC cohort was 1.8 times larger, so would be expected to identify a large predictor such as this. The finding of significance in the SPARCC cohort is lost if the BMD score is split into multiple levels (ie high, medium, low BMD score). For these reasons, the significance of the SPARCC BMD score is unlikely to be a true positive result.

Figure 3.26. AS susceptibility allelic risk scores (TASC cohort).

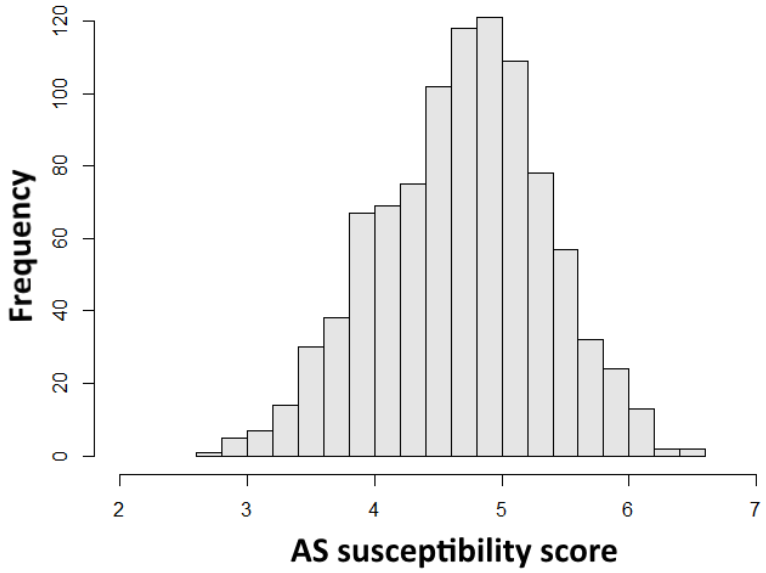


Figure 3.27. Observed cmSASSS vs AS risk score. There is no apparent relationship between the AS susceptibility based allelic risk score and observed cmSASSS. This observation holds across symptom duration. Attempts to fit a linear model (blue dotted line) find no significant linear relationship in any of these four symptom duration bin.

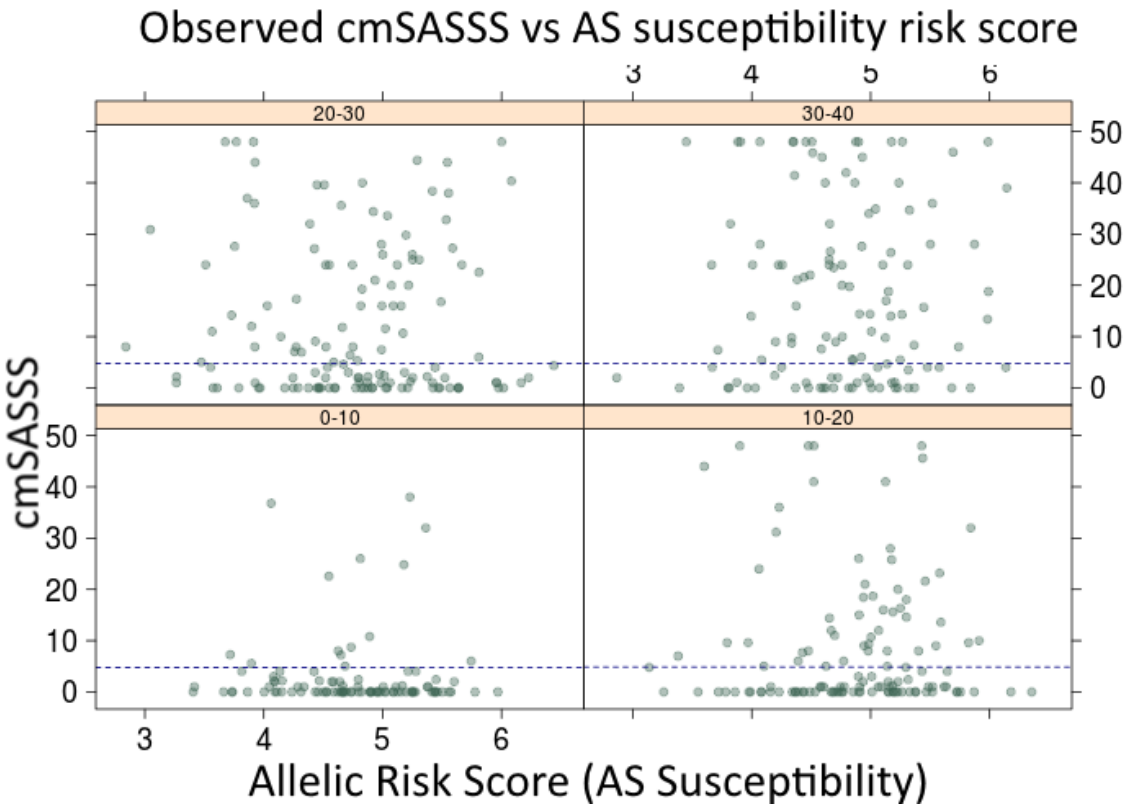


Table 3.16. Significance of allelic risk scores. P-value for signficance in the ZINB model. BMD was only significant in the SPARCC cohort.

Score	Cohort	Subjects	Score Significance
AS susceptibility	TASC	515	p > 0.59
AS susceptibility	SPARCC	201	p > 0.34
BMD	TASC	375	p > 0.27
BMD	SPARCC	210	p = 0.037
Height	TASC	374	p > 0.14
Height	SPARCC	210	p > 0.84
Nicotine dependence	TASC	387	p > 0.93
Nicotine dependence	SPARCC	260	p > 0.64

Apart from the BMD based score, there was no significant or suggestive association in either cohort, making a significant association on meta-analysis impossible. Additionally, meta-analysis would have decreased number of SNPs examined as they needed to be in common between the two cohorts.

The BMD based score was only of borderline significance (p=0.037) in the smaller SPARCC cohort. Even meta-analysis with an equal sized cohort would have yielded a p-value >0.1. Importantly, because this method uses many SNPs to create each score, significant associations in both cohorts would be required to justify further investigation.

3.4 Discussion

3.4.1 GWAS Power and Limitations

The power of this GWAS was determined by a number of factors, especially the expected heritability of risk SNPs, the number of subjects, and the number of SNPs genotyped. The association analysis also relies on adequate phenotype modelling.

The TASC and SPARCC cohorts had thousands of subjects with clinical covariates, mSASSS and genotypes, but when limited to subjects with all three forms of data, the number of subjects was reduced to a little over 1000.

The power of this study limited its ability to investigate low frequency variants, and common variants with effect sizes <0.2 of a standardised phenotype. From AS susceptibility GWAS and comparable quantitative GWAS of BMD, it would be expected that radiographic severity in AS is determined by numerous loci with a range of effect sizes. For example, the GEFOS2 study of BMD found dozens of SNPs with effect sizes <0.1 standard deviations and a few SNPs with effect sizes of 0.1 – 0.15 standard deviations (176). This indicates that bone metabolism is largely determined by SNPs with a small effect size.

GWAS have an intrinsic inability to identify specific types of causal variants. Variants with low heritability (ie low MAF or effect size) or in low LD with the nearest tag SNP are less significant on association testing. Additionally it is not clear if an association signal is from a rare causal variant with a large effect size or from a common variant with a small effect size (181). It could be argued that maximising LD between tag SNPs and causal variants by dense genotyping would reduce this problem and maximise power. However simulations by multiple authors suggest that increasing sample size has a much greater effect on power than increasing SNP coverage (149, 181). This GWAS had a limited ability to detect causal variants that have low heritability or are in low LD with tag SNPs.

Another factor that affects the power and accuracy of this study is the assumption that the ZINB model accurately represents the data. As discussed in Chapter 2, the ZINB model is the best available model and is reasonably able to predict cmSASSS, as demonstrated by the estimated model r^2 of 0.39. Nevertheless, models of non-parametric data rarely describe observed data perfectly. The challenge of modelling mSASSS has been addressed by previous authors either by longitudinal modelling that only examines change in score or by ZINB modelling (65, 79).

This GWAS was adequately powered to detect common variants with large effect sizes, while the phenotype was expected to be determined by variants with predominantly small effect sizes.

3.4.2 GWAS Findings

For the phenotype of radiographic severity, this GWAS identified no locus with significant association after Bonferroni correction for multiple testing ($p < 8.8 \times 10^{-9}$) or the arbitrary genome wide significance level of $p < 5 \times 10^{-8}$. Less than ten loci had a suggestive association ($p < 1 \times 10^{-5}$). None of these loci were identified by previous candidate gene studies of radiographic severity. None of the suggestive loci were identified in previous AS susceptibility studies, although the effect size for susceptibility and radiographic severity are likely to be different. The suggestive associations include loci with a range of functions, including some related to inflammation or bone metabolism.

TNF inhibitors are commonly used treatment for AS. Their ability to influence bone formation in AS is not currently clear, with one observational study finding a protective effect and multiple studies finding no effect. *TNFRSF1A* and *TNFRSF1B* are the two main soluble TNF receptors. A recent candidate gene study by Cortes *et al.* did not examine SNPs from either of these two loci, with the closest examined SNP to *TNFRSF1B* being 10 Mb away (116). The receptor *TNFRSF1A* has repeatedly been identified as an AS susceptibility locus (22, 36), and the study by Cortes *et al.* associated *RANK* (*TNFRSF11A)* with radiographic severity in AS. Polymorphisms in *TNFRSF1B* have been associated with response to infliximab in Crohn's Disease (182, 183) and may play a similar role in AS. The suggestive association of *TNFRSF1B* in this study reinforces the known involvement of the TNF pathway in the altered bone metabolism of AS. The exact relationship between TNF and excessive bone formation in AS is controversial. TNF inhibits the bone forming Wnt pathway by upregulating DKK-1 and sclerostin which inhibit the Wnt pathway (21, 26). A simple loss of function variant of *TNFRSF1B* is not likely to be the link between TNF and bone formation, as TNF inhibition in AS cases does not accelerate bone growth (184). However in a mouse model, dual blockade of DKK-1 and TNF promoted osteophyte formation (26). The nature of these signals at growing syndesmophytes is clearly complex, and may involve regulation by other pro-inflammatory cytokines. MRI studies suggest that osteoproliferation is more likely to occur at vertebral corners with persistent osteitis or where osteitis has resolved (13). Perhaps in these

circumstances the balance of pro-inflammatory and bone repair signals reduces the ability of TNF to inhibit Wnt via DKK-1. The involvement of the TNF pathway in bone formation is complex, but the fact that TNF is an innate pro-inflammatory cytokine ties inflammation with bone formation. The uncoupled theory of bone formation in AS postulates that bone formation is independent of inflammation. The uncoupled theory is supported by studies of mouse models that show no change in bone formation in the presence of TNF inhibitor (17). The association of two TNF receptors, *TNFRSF1A* and *TNFRSF1B,* and the RANK receptor *TNFRSF11A* with AS implies a close connection between inflammation and osteoproliferation. It provides further evidence that the innate inflammatory pathways of TNF and RANK influence bone formation in AS. This finding is an important contribution to our understanding of AS pathogenesis and implies that the TNF-Wnt pathway is responsible for osteoproliferation in AS.

NLRP7 and the NLRP family of NOD Like Receptors detect Pathogen Associated molecular Patterns (PAMPs) and regulate the inflammasome (185). *NLRP7* detects microbial acylated lipopeptides, and can activate or inhibit caspase-1 in inflammasomes (186, 187). Variants of NLRP family members have been associated with Crohn's disease, atopic dermatitis and vitiligo (188). The NLRP family has not previously been associated with AS susceptibility, despite multiple GWAS and a candidate SNP study that examined three SNPs in *NLRP3* (189). Mouse models of *NLRP3* overactivity demonstrate excessive inflammasome activation, dramatically increased IL-1β production and multi-organ inflammation. The mice develop inflammatory arthritis and a Th17 predominant response (190, 191). The involvement of IL-1 in AS was previously suggested by the AS Immunochip study, which found *IL-1R1* and *IL1-R2* to have suggestive associations, $p < 5 \times 10^{-7}$ (22). IL-1β is linked to bone formation by its induction of the Th17 response, which is known to promote bone formation at entheses (25). Excessive activity of IL-1β in AS is also suggested by the strong association signal at *VRK2*, which modulates expression of IL-1β and other cytokines (179). Hence association of the NLRP associated inflammasome and *VRK2* in AS provide a theoretical mechanism for bone formation via IL-1β and IL-17/23.

NLRP inflammasomes also influence and are influenced by the gut microbiome. Presence of commensal bacteria amplifies the IL-1β dependent Th17 response (192). *NLRP6* regulates gut mucosal immunity and microflora. *NLRP6* knockout mice develop inflammatory colitis and dysbiosis of gut microbiota, due to impaired clearance of enteric pathogens (193, 194). These mice have a higher abundance of *Prevotellaceae*, gram negative bacteria that elicit a strong inflammatory response from the gut mucosa (193). When transferred from *NLRP6* knockout mice to healthy wild-type mice, *Prevotellaceae* induce colitis. A recent study by Costello *et al.* found that human AS cases also have a higher abundance of *Prevotellaceae* compared to healthy controls (195). Loss of function variants in NLRPs causing impaired clearance of *Prevotellaceae* is a possible mechanism to explain the change in abundance and consequent colitis in AS. Chronic colitis with inflammasome activation would produce IL-1β, causing an IL-17/23 response which would promote chronic bone formation. Variation in NLRP function is a potential mechanism of innate immune system overactivity in AS. It provides a theoretical template for understanding the mechanism for excessive acute inflammation, intestinal dysbiosis, colitis, the Th17 response and bone formation.

Previous research has not identified the mechanisms that trigger the innate immune response in AS. A candidate SNP study suggested *CARD8*, and not *NLRP3* were associated with AS susceptibility. This study only compared one SNP in *CARD8* and three SNPs in *NLRP3* (189). AS susceptibility studies have identified involvement of inflammasome components (eg *CARD9*) but not genes in the pathways that activate the inflammasome (22, 36). NLRPs may activate the innate immune system in AS. Deleterious *NLRP7* variants may promote to inflammasome overactivity, promoting a Th17 phenotype and changes to the intestinal microbiota, resulting in colitis and arthritis. Once intestinal dysbiosis is established, the inflammatory phenotype would be maintained by aberrant host-microbiota interactions from *NLRP7*. *NLRP7* therefore provides another potential pathway for innate immune system activation and chronic inflammation in AS.

Confirmation of the involvement of the NLRP family in AS pathogenesis would strengthen the notion that AS involves aberrant activation of innate inflammation by pathways similar

to IBD. Modulation of these innate pathways should influence severity of inflammation and osteoproliferation. For example in AS and Crohn's disease, TNF inhibitors are effective at reducing inflammation by blocking TNF, an innate pro-inflammatory signal. This GWAS provides further evidence that innate pathways drive inflammation and bone formation in AS.

FGF4 is involved in cartilage and bone metabolism and wound repair. It regulates skeletal formation and chondrogenesis during development and is also expressed in adult mice (196-199). Gene therapy with FGF4 has also been shown to improve soft tissue wound repair and to stimulate fibroblast proliferation (200). The closely related *FGF2* promotes endochondral ossification in fracture repair, and acts synergistically with *BMP2* to increase ossification by periosteal cells via canonical Wnt signalling (201-203) . Case reports of patients with craniosynostosis (premature ossification of cranial sutures), have found duplication of the 11q13.3 region containing *FGF3* and *FGF4* (204). The process of syndesmophyte formation in AS is believed to involve erosion of cartilage and bone, which is replaced by cartilaginous repair tissue which later ossifies (11). Similar to craniosynostosis, osteoproliferation in AS may be caused by overactive FGF pathways, initiated to repair erosions. This would provide a link between the events of bone erosion and bone formation in AS, and would support the coupled hypothesis. The association between the skeletal growth factor *FGF4* with syndesmophyte formation should be investigated as it may provide the link that couples erosion to osteoproliferation in AS.

A suggestive association exists at *BMP2*, a gene that stimulates osteogenesis, skeletal homeostasis and fracture repair (205). The suggestive association was only present at markers with MAF < 0.1. The association was present at low levels of significance in common genotyped SNPs from both cohorts. The previous candidate gene study by Cortes *et al.* (116) included ten SNPs within *BMP2*, three of which had $p<0.05$ for either cervical, lumbar or total mSASSS. No SNP in *BMP2, BMP3, BMP4, BMP6, BMPR2, BMPR1A* or *BMPR1B* met experiment wide significance. Together these features make the significance of this locus uncertain. Similar to *FGF4*, overactivity of *BMP2* would be expected to cause excessive bone formation following bone injury or resorption from

normal bone turnover. Aberrant *BMP2* activity in AS would also provide a link between erosion and osteoproliferation. The fact that this study as well as the study by Cortes *et al.* found some degree of association at *BMP2* could be attributed to overlapping cohorts and the association of low frequency variants only. Despite this uncertainty, the known relevance of *BMP2* to bone metabolism makes it a target worthy of investigation.

The suggestive associations at *OFCC1*, *VRK2*, 3p24.1, 4q21.23 are at loci with poorly defined functions. There is also a suggestive association at *HTR5A* which has no evident involvement in spondyloarthropathy, inflammation or osteoproliferation. *OFCC1* is believed to cause an orofacial cleft malformation but its role in skeletal development and metabolism has not been ascertained. *VRK2*, a protein kinase, has been linked to transcription of IL-1 β and is expressed in monocytes and neutrophils (179). Its activity could therefore influence a range of autoimmune conditions. It has not been identified as significant by association studies of any IL-1β related autoimmune disease, making its significance tied to the involvement of other IL-1β related signals. The suggestive association in an intergenic region at 3p24.1 could be from a regulatory sequence for a predicted transcript 216 bp in the 3' direction or for the *EOMES* which is 255kb in the 5' direction. *EOMES* is a transcription factor implicated in AS and Multiple Sclerosis that is associated with variation in $CD8^+$ lymphocyte counts (22). If the association signal is from *EOMES*, it would be the only suggestive association with any element of acquired immunity. The suggestive association at 4q21.23 may be from a predicted lincRNA expressed in muscle, leucocytes, fibroblasts and chrondrocytes. It may have regulatory functions related to wound repair, or metabolism of cartilage or bone.

No loci with suggestive associations in this study were identified by previous candidate gene studies, as apart from *BMP2* these loci were not investigated. Candidate causal variants identified by previous studies (eg *PTGS1*, HLA alleles, *NELL1*, *LMP2*) were not found to be associated by this study. These candidate gene studies were vastly different analyses in terms of genotyping, imputation and phenotype modelling. It could be argued that this study, as the first genome wide study, and the first study to attempt complex

phenotype modelling has used the best approach to date. The suggestive associations found by this GWAS require confirmation by future candidate gene studies or GWAS.

GWAS suggests that genes related to the innate immune system, microbial interaction and bone repair are related to bone formation in AS. These genes provide clues about the link between bone erosion and bone formation in AS, and the initiation and prolongation of the innate immune response and the Th17 response. Future study should investigate the function of these genes and their pathways in AS cases, and compare the microbiome of risk variants.

3.4.3 HLA Association

This was the first association study to compare the entire HLA region with the phenotype of radiographic severity. Similar to the vast majority of previous studies, the phenotype was found to be independent of *HLA-B27* status. No other HLA allele or amino acid polymorphism was significantly associated with the phenotype.

A previous candidate gene study by Ward *et al*. (71), identified numerous HLA alleles as predictors of radiographic severity. In the combined TASC and SPARCC cohort, no HLA alleles listed by the previous study were significant predictors. All candidate alleles identified by the study were low frequency variants (MAF < 0.1) with six out of eight being rare variants (MAF <0.05).

Considering the results of the previous study by Ward *et al.* and this larger entire HLA analysis, there is no evidence to suggest that any polymorphism within HLA strongly contributes to radiographic severity. This finding contrasts with the established role of HLA alleles in disease susceptibility. AS susceptibility causal variants in the HLA probably contribute to inflammation without affecting syndesmophyte formation over time. Osteoproliferation is probably more strongly influenced by other pathways, such as innate inflammatory or bone repair pathways than acquired immunity initiated by HLA signalling. The lack of strong association between HLA alleles and radiographic severity is an indication that osteoproliferation in AS is driven by factors outside HLA related pathways.

3.4.5 Genome Wide Complex Trait Analysis

GCTA demonstrates that variation in radiographic severity is strongly associated with genetic variation across the genome. Similar to other diseases, SNPs that do not pass the stringent genome wide significance level do make a significant contribution to the phenotype (45, 206). Apart from Chromosomes 5 and 3, the per chromosome analyses were not significant, likely due to the limited number of examined SNPs on each chromosome and therefore the proportion of genetic variance captured by the genotypes available.

This is the first study using genome wide data to demonstrate the heritability of radiographic severity in AS. It conservatively estimates the heritability to be at least 0.33, meaning that at least this proportion of variation in the phenotype is from genetic factors. Previous estimates of determinants of radiographic severity have attributed at least 23-38% of phenotypic variation to non-genetic factors (Table 2.12, (3)), and 56% to genetic factors (6). Given the limitations of this GCTA, and the limited contribution of non-genetic factors identified by many previous studies, heritability probably contributes more than non-genetic factors to the variability of this phenotype. The true value of broad sense heritability is likely to approach two-thirds. The simpler BASRI score of AS severity was estimated to have a broad sense heritability of 0.56 by a pedigree study (6). It could be argued that this BASRI value, calculated by a pedigree study, is an overestimate due to the shared environment of family members. A shared environment should not influence this GCTA as it compares specific genetic variants present in a population of 1020 subjects (with related individuals removed) rather than a small number of families. An improved GCTA estimate could be obtained with a sample size of 10 000 subjects, an approach used by prominent GCTA studies of quantitative phenotypes (47, 168, 207).

The GCTA findings are promising for the role of genetics in the clinical management of AS. They suggest causal genetic variants are present and measurable. Once identified, a prognostic score could be created to guide long term pharmacotherapy and aid patients in

life decisions. The results also suggest that multiple loci with small effect sizes influence radiographic severity. These are important drug targets, even if their effect size is small. In the case of AS susceptibility, pathological relevance is not proportionate to locus effect size. The TNF pathway is an excellent therapeutic target, yet *TNFRSF1A* has OR 1.13 compared to *HLA-B27* with OR 60 (22). The GCTA findings suggest unidentified causal variants are present, which should be investigated for clinical use.

A particularly interesting finding of GCTA was that almost half of the whole genome V_G/V_P was present in Chromosome 5. This finding was based on a large degree of variation in 18 070 SNPs examined from Chromosome 5 and was highly significant. There were no suggestive associations from Chromosome 5 on the GWAS, with no SNP having $p < 1 \times 10^{-5}$. AS susceptibility loci present on chromosome five include *ERAP1, ERAP2, LNPEP, PTGER4, IL-12B* and *IL-7R*. None of these loci have been directly linked to bone formation, although *PTGER4* encodes a prostaglandin receptor, and *IL-12B* encodes a subunit of the IL-23 receptor, which promotes bony spur formation (25). The variation detected in this chromosome may be from these loci or from unidentified AS related loci. The presence of this strong signal in this relatively small GCTA sample suggests this chromosome is deeply involved in AS pathogenesis and should be examined for variants associated with osteoproliferation.

Despite thorough genotyping of the HLA region with the Illumina Immunochip platform, the variation in Chromosome 6, which includes the HLA, was not found to be positively associated with the phenotype. This adds to the findings of the HLA association testing, which do not demonstrate any contribution of the HLA region.

GCTA can only compare the genetic variation presented to the software. In this case, almost 400 000 SNPs were examined in over 1000 subjects. Other GCTA studies commonly use over 5 000 subjects, and often use over 10 000 (47, 169). The SNPs examined in this analysis are only a fraction of the genetic variation present in each subject. An ideal analysis would include millions of genotyped SNPs from each subject,

with all causal variants tagged. Because some causal genetic variation was probably not tagged in this study, the narrow sense heritability is probably higher than the GCTA estimate. Additionally GCTA uses an additive genetic model so cannot assess heritability from other genotype-phenotype relationships, such as dominance, gene-gene interactions and gene-environment interactions. For these reasons the estimate of narrow sense heritability by GCTA will always be lower than the true broad sense heritability. Given the limitation of this analysis, the heritability of AS radiographic severity is at least 0.33.

GCTA demonstrates that the phenotype of radiographic severity in AS is determined to a large degree by genetic variants that did not reach genome wide significance on GWAS. The phenotype is polygenic and is probably influenced by genetic variants more than environmental factors. Further characterisation of these genetic variants would improve clinical management.

3.4.4 Allelic Risk Score

To date no loci associated with AS susceptibility has been associated with AS radiographic severity. The most apparent cause of failing to find loci with association would be an insufficient sample size to detect the effect size. The allelic risk score analysis provides some ability to find a relationship between multiple loci with small effect sizes and the phenotype. Once causal variants have been established, it can also be used for risk prediction. The allelic risk score approach is most useful for polygenic traits with small effect sizes that are not highlighted by association testing of multiple independent loci by GWAS.

An AS susceptibility allelic risk score assessed the contribution of forty AS susceptibility loci to radiographic severity. Similar to the genome wide and HLA association testing, this polygenic analysis did not find association between loci known to cause AS susceptibility and radiographic severity. *HLA-B27* had to be removed from the score as its weighting (OR 60) made other loci irrelevant. The score created from other AS susceptibility loci was not predictive of radiographic severity. If inflammation is closely linked to bone formation in AS, inflammation related AS susceptibility loci might also be determinants of radiographic severity. The allelic risk score analysis provided no evidence to support this. The AS susceptibility allelic risk score did not identify the expected relationship between AS susceptibility loci and osteoproliferation.

Genes involved in bone metabolism are candidates for genetic determinants of syndesmophyte formation. Individuals with bone maintenance or repair pathways that are prone to more bone deposition would presumably have a higher risk of syndesmophyte formation. This analysis found no consistent evidence of the involvement of osteoporosis related genes in osteoproliferation. If this finding were replicated on a larger scale it would be an indication that the processes that affect BMD over a lifetime differ to the processes that contribute to syndesmophyte formation. This is unlikely given the involvement of inflammatory mediators such as RANK in both BMD and AS, and the connection between TNF and the bone forming Wnt pathway by *DKK-1* (27, 116, 176). The BMD based allelic

risk score did not identify the expected relationship between bone metabolism genes and osteoproliferation in AS.

SNPs associated with human height were examined as previous studies have suggested a connection between bone growth loci and human height. Analysis of height GWAS have identified numerous height associated SNPs in or near known bone or skeletal genes, including *BMP3, BMP6, WNT6, WNT3A, WNT9A, MEF2C* and *STARD3NL* (174). The height based allelic risk score was not found to be predictive of radiographic severity. The score may be misdirected in a number of ways. It may focus on loci responsible for bone formation early in development, or may include alleles that predispose to bone formation in healthy individuals but not in AS. Importantly the phenotype of human height is not the same as the osteoproliferative phenotype. Human height is based on many factors, including the length of long bones. Osteoproliferation in AS is more likely to involve genes related to bone mineralisation, bone repair and endochondral ossification. Osteoproliferation related alleles in the height score may be diluted beyond significance by alleles related to height only. Despite the relationship between bone metabolism and human height, the role of height loci in the dysfunctional bone metabolism of AS remains to be established.

Cigarette smoking is an established predictor of radiographic severity in AS (71, 79). Individuals with an increased risk of nicotine dependence are expected to be at risk of radiographic progression. This analysis found no association between the nicotine dependence allelic risk score and radiographic severity. Alleles associated with nicotine dependence are an indirect measure of smoking, which is influenced by many environmental and cognitive factors. As the relationship between smoking and radiographic severity is strong, yet the risk score has no relationship, cognitive and environmental factors are probably the main predictors of smoking in this cohort.

The allelic risk score analysis is a simple system with multiple limitations. The most important requirement is that the SNPs included in the score must contain causal variants, or variants in LD with causal variants, of the phenotype being tested. Another fundamental assumption of the weighted allelic risk score is that each loci's risk score effect size is

similar to its radiographic severity effect size. It has been repeatedly established that the effect size of *HLA-B27* on AS susceptibility is very different to its effect size on radiographic severity. It is therefore reasonable to assume that the effect sizes for other loci may differ greatly in their contribution to susceptibility and radiographic severity. Loci included in this analysis may contribute to AS radiographic severity but if weighted incorrectly the effect may not be visible. An alternative approach is to weight all loci equally, but again if these equal weights differ greatly from the actual effect size of each locus the relationship will be lost. Similarly, if only a small fraction of loci included in the score are causal variants, the association will be diluted beyond significance. Another shortcoming of this technique is that it estimates the risk from each locus by including only the risk from the lead SNP, not from all risk SNPs. The allelic risk score is useful tool to investigate polygenic traits, but has major limitations, which is the likely reason it was unable to link these related phenotypes.

The allelic risk score analysis was performed in an attempt to capture the possible contribution of many loci with small effect sizes. Using four scores based on phenotypes related to AS, no connection was established.

3.5 Conclusion

This study was the first genome wide comparison of genotype and radiographic severity in AS. It was the first study to compare all HLA types with radiographic severity and to investigate the contribution of common genetic variants with low effect sizes to this phenotype.

The GCTA study indicates that at least one third of phenotype heritability is from genetic variation. This is the first genetic marker based study to confirm that a large proportion of AS radiographic severity is determined by genetic factors. The conservative GCTA approach estimated phenotype heritability to be 33%, but given the lack of other predictors of radiographic severity, and the strong genetic component of AS susceptibility, the expected narrow sense heritability of AS is probably significantly higher than this. The true value is probably at least as high as the estimated broad sense heritability of radiographic severity in AS using the BASRI score, 56%. A broad sense heritability for this phenotype of 60% would be consistent with the phenotype modelling (Chapter 2) which attributed up to 39% of phenotypic variance to environmental factors and symptom duration. The source of heritability identified by GCTA is likely to be common variants of low effect size rather than low frequency variants, as these are not captured as well by the genotyping approaches used. GCTA confirms the heritability of this phenotype, but the relative contributions of common and rare variants will need to be elucidated by large GWAS.

The main clinical implication of this heritability finding is that AS severity appears to be more strongly influenced by genetic factors than environmental factors. Phenotype modelling (Table 2.12) found <4% of variation was attributable to smoking, which is far smaller than that attributed to risk variants by GCTA. Therefore, disease modification in AS should focus on treating aberrant cell biology rather than environmental exposures of limited influence.

This study suggested differing roles of acquired and innate immunity in AS pathogenesis. Highly accurate HLA imputation of more than 1000 cases did not associate any HLA type with radiographic severity. Similarly GCTA did not find evidence that variation in chromosome 6 was significantly associated with the phenotype. If replicated in a larger cohort, this would mean that HLA variation has little or no influence on bone formation in AS which suggests that acquired immunity does not influence it either. Instead the GWAS findings suggest innate inflammatory pathways including TNF and inflammasomes are involved in osteoproliferation. If confirmed, this study's findings would suggest that while acquired immunity is a strong contributor to disease susceptibility, progressive bone formation in the spine is probably influenced by other factors, including innate inflammation.

The GWAS findings also contribute to the discussion of mechanisms of bone formation in AS. The GWAS results suggest an association between the innate immune system involvement and bone formation in AS. *NLRP7* containing inflammasomes and the TNF pathway and were associated with mSASSS. This links acute innate inflammation with bone formation. The anatomical location of the innate inflammation that drives bone formation is not clear. The *NLRP* family are known to be act as microbial sensors and regulators within leucocytes in gut mucosa. Activity of this family of genes at entheses or during bone repair has not been described. *NLRP7* variants may simply stimulate and maintain the Th17 pathway in the gut, causing colitis and systemic inflammation. The link between *NLRP7* and Th17 has not been established, but overactivity of the closely related *NLRP3* stimulates a Th17 response (191). An intestinal Th17 response has been shown to trigger inflammatory arthritis with bony spur formation in a spondyloarthritis mouse model (208). Thus the *NLRP* family are a potential link between innate inflammation and Th17 induced bone formation in AS.

Another innate inflammatory pathway, TNF, is known to be active in both colitis and osteitis. The results of this GWAS suggest that the TNF-Wnt pathway contributes to bone formation in AS. The role of TNF as a regulator of bone formation is controversial given the conflicting findings that TNF inhibition does not appear to increase syndesmophyte formation in cases, may inhibit bone formation in cases, and did not affect bone formation

a mouse model (17, 79, 184). The mechanism by which the TNF-Wnt pathway promotes bone formation has not been established, but this GWAS finding suggests that TNF does influence bone formation in AS. The suggestive association of two innate inflammatory pathways (NLRP and TNF) implies inflammation is tightly linked to bone formation in AS. These suggestive associations require replication and confirmation of effect size. Activity of NLRPs could be compared between AS cases with and without active bone formation.

Two bone metabolism loci, *FGF4* and *BMP2* had suggestive associations in this GWAS. These genes are involved in wound repair, and may be activated by bone erosion in AS. Overactivity of these genes would cause excessive bone formation in response to erosion. These loci could be investigated by examining their level of expression at sites of osteoproliferation or by known down studies.

This GWAS did not associate known AS susceptibility loci with the phenotype. This analysis used a dramatically different approach to susceptibility studies and consequently highlighted different loci that contribute to the related phenotype of radiographic severity. The method of quantitative phenotype analysis and the differing effect sizes that AS susceptibility loci have for susceptibility and severity are the likely reason why this GWAS identified new AS related loci in similar pathways to the AS susceptibility GWAS.

The allelic risk score analysis did not find evidence that AS susceptibility loci, bone metabolism loci, nicotine dependence loci or height loci contribute to the phenotype of radiographic severity. While the allelic risk score method provides only a snapshot of each subject's genome, it was an attempt to create a simple clinically applicable prognostic score. Presumably the causal variants are in loci related to the risk score loci, but were not tagged by SNPs used in the risk scores.

This study had major acknowledged limitations which affected the statistical significance of the results. The sample size and phenotype modelling were the best available to the researchers but both limited the study's ability to detect causal variants. The GCTA

estimate may have been limited by the small sample size. Small GCTA studies include 1 000 to 3 000 subjects, but prominent GCTA studies have sample sizes of 10 000 subjects (47, 168, 169, 207). For the GWAS, HLA association testing and the allelic risk score analysis, the sample size only provided adequate power to detect common variants with a large effect size. GWAS with low power have limited ability to detect causal variants in low LD with tag SNPs. The allelic risk score analysis which relies on causal variants being in high LD with tag SNPs was significantly hindered by the low power. The complex phenotype of syndesmophyte formation and ankylosis was difficult to model using a cross-sectional approach. A complex ZINB model was used as too few subjects had multiple radiographs to allow longitudinal analysis. The adequacy of the ZINB modelling is difficult to assess. Diagnostic model tests and sensitivity testing found the model to be adequate and the use of ZINB models by previous authors to model radiographic scores, including AS, is an indication that it is the best approach available. An intrinsic limitation of the NB and ZINB models was the propensity to underestimate mSASSS in subjects with severe disease, resulting in the increased weighting of subjects with ankylosis. This effect was clearly seen in the SPARCC cohort but was only a minor effect in the much larger TASC cohort. The scoring of films by multiple readers is another potential source of error. Minor inter-rater variation may have been present as most films were not scored by multiple readers. The effect of inter-rater variation was probably minimal as the most subjective element of mSASSS, scores for erosions, sclerosis and squaring, were removed. Inter-rater variation would be minimised if films were scored by multiple readers or a single reader, or if longitudinal scores were analysed. Longitudinal modelling with more extensive environmental exposure data, including disease activity, occupational and treatment data would enable more precise phenotype modelling. Patient recall of disease symptom onset is likely to be a cause of error, which would be avoided using longitudinal studies. Obtaining an adequate number of paired radiographs with a sufficient interval between them is challenging. Further, such data would be subject to reader error with two sets of measurements, not just the one as is the case in cohort studies. Other unidentified predictors of cross-sectional severity may also exist. A larger cohort would also improve modelling, and if more than 3000 subjects were included, the GWAS power would improve dramatically, providing increased certainty regarding significant loci and improved

detection of low frequency variants. This study had major limitations which should be considered by the next study into radiographic severity.

The findings of all the genetic investigations suggest that genetic factors make a significant contribution to AS radiographic severity, probably more than all environmental factors combined. No common risk alleles with large effect sizes were identified, but GCTA suggests that common variants with small effect sizes make a significant contribution, with at least one third of phenotype variation attributable to genetic variation. HLA alleles and acquired immunity in general do not have a major influence bone formation in AS. While HLA alleles are the main determinants of AS susceptibility, they do not appear to exert significant influence on bone formation over years. Causal variants appear to be present in innate immune system pathways which suggests a close link with bone formation. A possible new link between the Th17 response in the gut mucosa and bone formation in AS has been suggested. These findings require replication in a larger cohort, but highlight areas of cell biology which with further characterisation may provide valuable therapeutic targets. Future research should focus on identifying: common causal variants, the link between the suggestive GWAS associations and bone formation, and the potential role of the NLRP family and *Prevotellaceae* in the IL-17/23 response of AS.

Thesis Conclusion

Ankylosing Spondylitis is a function limiting condition with few effective pharmacological agents. Currently clinicians have a limited ability to predict or prevent future ankylosis. Identification of modifiable environmental or biological factors that influence ankylosis is an important goal in the management of this disease.

This study identified and quantified the major known contributors to radiographic severity - gender, smoking and disease duration. Smoking duration was found to be a more influential predictor of the phenotype than pack year history. Common variants with small effect sizes were found to contribute to at least 30% of phenotype variation. No signal was identified to link known contributors to AS susceptibility, including the MHC region, with radiographic severity. Known AS susceptibility, BMD and height loci were not associated with severity, meaning that their effect size for severity was small or zero. Genes in the same pathways as AS susceptibility loci, including the TNF pathway, were associated with radiographic severity. A new suggestive association linking bone formation to immune-microbiota interactions in the gut was made, which may explain observed patterns of microbiota in human AS cases. Genetic associations also suggested that bone metabolism and inflammation related loci are related to the phenotype, although findings were not definitive. Reasons for the study failing to identify definite genetic associations include its low power, errors in phenotype modelling from to a cross-sectional approach and lack of correction for other non-genetic predictors of severity such as medication.

As the first comprehensive study of the relationship between genotype and radiographic severity, this study confirms that genetic factors make a significant contribution. Multiple epidemiological studies have searched for non-genetic determinants, yet have not been able to attribute more than one third of variation to specific factors. The genetic findings indicate that common variants, the innate immune system, bone metabolism and perhaps the microbiome all affect the phenotype. The male gender is by far the largest genetic determinant of severity. These contributors should all be investigated with much larger

genetic association studies as well as expression and functional profiling of the new candidate genes in AS cases. Once replicated, significant genetic determinants of radiographic severity would make attractive targets for therapeutic intervention.

Appendix

A.1 Symptom Score Details

Visual Analogue Scale Stiffness

- How much stiffness has the patient had because of the illness in the past week?

Early Morning Stiffness

- If the patient is stiff in the morning, about how long does the stiffness last (in minutes)?

Exercise number per week

- How many times does patient exercise in a typical week?

Exercise minutes per week

- How long does each exercise session typically last?

Back Exercises days per week

- Does patient do any back stretching or strengthening exercises?

A.2 R Output for NB model of TASC contracted mSASSS

```
Call:  glm.nb(formula = round_any(Tot_012, 1) ~ DISDUR + GENDER, data = full.xr,
link = log)
Deviance Residuals:
 Min       1Q   Median       3Q      Max
-1.8513  -1.3488  -0.5946   0.1888   3.0889

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.314014   0.110914   2.831  0.00464 **
DISDUR      0.048185   0.003146  15.317  < 2e-16 ***
GENDERM     1.020703   0.105119   9.710  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(0.379) family taken to be 1)

Null deviance: 1775.1  on 1391  degrees of freedom
Residual deviance: 1481.5  on 1389  degrees of freedom
AIC: 8381.3

Number of Fisher Scoring iterations: 1
```

```
              Theta:  0.3790
          Std. Err.:  0.0171

 2 x log-likelihood:  -8373.3460
```

Calculation of Dispersion Parameters for NB model

φ = NB model deviance / residual degrees of freedom in NB model = 1481.465/1389 = 1.067

$\alpha = 1 / \text{theta}_{\text{NB model}} = 1 / 0.379 = 2.639$

A.3 ZINB R OUTPUT of chosen model

This model is listed as 'Model A' from Table 2.12.

```
Call:
zeroinfl(formula = Tot_012.success ~ GENDER + READER + DISDUR *
smoking.duration,
    data = full.xr.smkdur, dist = "negbin", EM = FALSE)

Pearson residuals:
    Min      1Q  Median      3Q     Max
-0.9403 -0.5927 -0.3550  0.3161 10.0629

Count model coefficients (negbin with log link):
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)             1.2808113  0.2014353   6.358 2.04e-10 ***
GENDERM                 0.5751680  0.1089477   5.279 1.30e-07 ***
READER2                 0.1224244  0.1574228   0.778  0.43676
READER3                 0.3590427  0.2142943   1.675  0.09384 .
READER4                 0.2191919  0.1773103   1.236  0.21638
READER5                 0.3070046  0.2068312   1.484  0.13772
READER6                 0.5375261  0.3456145   1.555  0.11988
READER7                0.1522920  0.1886301   0.807  0.41946
DISDUR                  0.0276956  0.0039376   7.034 2.01e-12 ***
smoking.duration        0.0230824  0.0072838   3.169  0.00153 **
DISDUR:smoking.duration -0.0003600  0.0001798  -2.002  0.04524 *
Log(theta)             -0.0379427  0.0771288  -0.492  0.62276

Zero-inflation model coefficients (binomial with logit link):
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)             1.1264424  0.4499395   2.504   0.0123 *
GENDERM                -1.0244977  0.2011589  -5.093 3.52e-07 ***
READER2                 1.0309194  0.4328975   2.381   0.0172 *
READER3                 1.1706588  0.5519414   2.121   0.0339 *
READER4                 0.7876221  0.4974656   1.583   0.1134
READER5                 0.8532399  0.6534593   1.306   0.1916
READER6                 0.0423468  0.9628321   0.044   0.9649
READER7                 0.1915973  0.5020500   0.382   0.7027
```

```
DISDUR                  -0.1001037  0.0100910  -9.920  < 2e-16 ***
smoking.duration        -0.0978515  0.0193117  -5.067 4.04e-07 ***
DISDUR:smoking.duration  0.0023821  0.0006067   3.926 8.62e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Theta = 0.9628
Number of iterations in BFGS optimization: 29
Log-likelihood: -3214 on 23 Df
```

A.4 Goodness-of-fit plots for NB model (glm.nb function in *MASS* package)

Standard goodness-of-fit tests to assess homogeneity, independence of X and Y, outliers and influential observations. Outlier subjects 1345, 1137 and 955 were checked for invalid observations, and none were found. The histogram and QQ plot of residuals suggest an excess of smaller residuals than expected. The plot of NB model residual vs observed values demonstrates that the model underestimates higher scores, and overestimates lower scores (causing the negative residual). A zero-inflated model should address the latter issue. NB model fitted values range up to 141, about three times the maximum cmSASSS value of 48. The abrupt upper and lower bounds of datapoints on the residual vs fitted plot are due to observed cmSASSS ending abruptly at 0 and 48.

Standard goodness-of-fit tests to assess normality of residuals, homogeneity, independence of X and Y, outliers and influential observations. Outlier subjects 1345, 1137 and 955 were checked for invalid observations, and none were found. The histogram and QQ plot of residuals suggest an excess of smaller residuals than expected. The plot of NB model residual vs observed values demonstrates that the model underestimates higher scores, and overestimates lower scores (causing the negative residual). A zero-inflated model should address the latter issue. NB model fitted values range up to 141, about three times the maximum cmSASSS value of 48. The abrupt upper and lower bounds of datapoints on the residual vs fitted plot are due to observed cmSASSS ending abruptly at 0 and 48.

Figure A.4.1 Pearson residual plot vs symptom duration plot. This demonstrates independence of the residuals from symptom duration (the model predictor with the largest r^2).

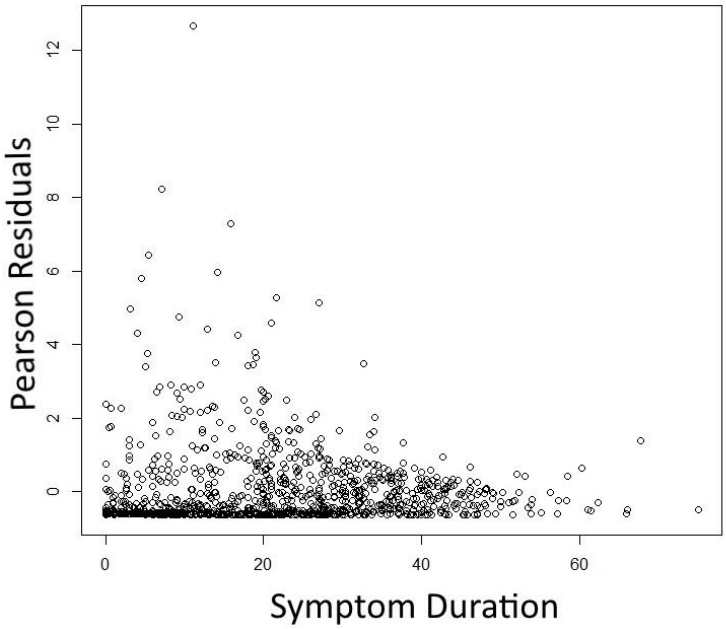


Figure A.4.2 Pearson residual plot vs fitted plot. This demonstrates adequate homogeneity and is distributed as expected (See A.4.3 for expected distribution).

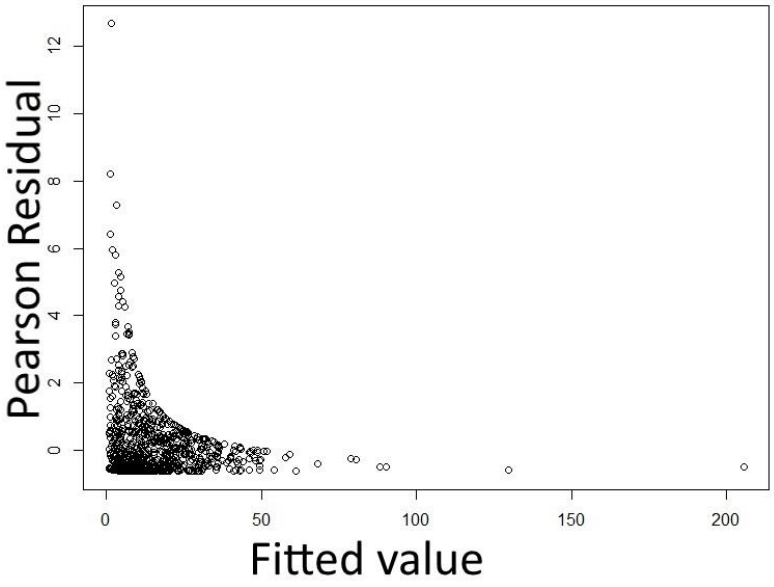


Figure A.4.3 Typical Pearson residual vs fitted plot for a ZINB model. Reproduced from Hilbe (Figure 9.3) with permission from Cambridge University Press (113). Pearson residuals of Generalised Linear Models are not expected to be normally distributed. As Zuur et al. note (p230) "*However, it should be noted that we are not looking for normality from the Pearson or deviance residuals. It is all about lack of fit and looking for patterns in the deviance or Pearson residuals*."

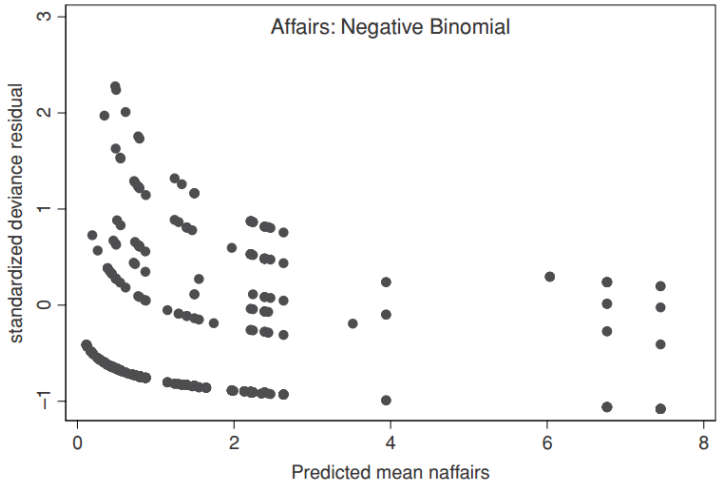
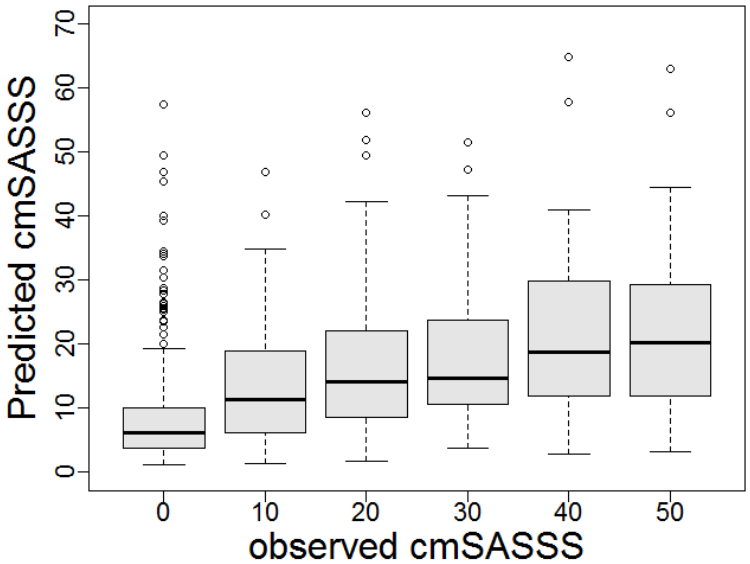


Figure A.4.4 Predicted vs observed boxplot. This demonstrates that predicted cmSASSS increases as observed mSASSS increases. The size of predicted values are proportionately lower than the observed values. For example the median predicted value for observed cmSASSS of 20 is 14.



A.5 Goodness-of-fit plots for ZINB model (zeroinfl() function, *pscl* package)

Compared to the goodness-of-fit plots for NB, the final ZINB model has more normally distributed raw residuals which have a more random pattern when plotted against symptom duration. The ZINB model also has smaller raw residuals which remain in the 0-48 range of the cmSASSS scale. The Pearson residual vs fitted plot is as expected (variance decreases with higher fitted values), and the predicted value increases as the observed value increases. Covariates such as symptom duration have no remaining relationship with the Pearson residuals. A QQ plot of Pearson residuals with an expected residual envelope confirmed an adequate model fit and robustness against influential outliers.

Notably predicted vs observed plots demonstrate predicted values are lower than observed, especially at higher observed values. This may be due to the fact that the negative binomial distribution does not include a peak for completely ankylosed individuals with a maximum mSASSS (as demonstrated in Figure 2.11).

Equation A.5.1. *Definition of Pearson Residuals*. Pearson residuals are calculated by dividing observed minus fitted values by the square root of the variance function. Notably the NB distribution has a large variance at zero which decreases dramatically as x increases. Small Pearson residuals indicate that the difference between observed and fitted values is small compared to the variance of the NB distribution at this point.

$$r_i^P = \frac{y_i - \mu_i}{\sqrt{Var(\mu_i)}}$$

Figure A.5.1. Pearson residual plot vs fitted values. Figure A - The distribution of residuals is uneven with many small Pearson residuals, including many less than zero. Figure B is the same plot as Figure A, with observed cmSASSS values <=5 in green, and those >5 in red. This demonstrates that the many low Pearson residuals are produced from the hundreds of subjects with observed cmSASSS <5. Following modeling with fixed effects and the binomial and NB distributions in the ZINB model, the predicted (fitted) values are less clumped under 5 cmSASSS points. Observations with a high fitted value have a higher Pearson residual, while many observations with a low fitted value have a small, negative Pearson residual. Negative Pearson residuals indicate that the expected (fitted) value is higher than the observed. These occur because the ZINB model underestimates the number of zeros (and low values) as the count model adds fixed effects for gender and Scorer. For example, all males are given a fixed effect of 0.6 units, so a male with an observed cmSASSS of zero and a symptom duration of zero would have a predicted cmSASSS score slightly higher than zero, producing a negative residual. Small Pearson residuals indicate a small ratio between the difference in observed and fitted values compared to the variance function. The variance function is greatest approaching zero, so low variance of the residuals at this point will generate small Pearson residuals. Thus the significant number of small and negative Pearson residuals is a result of many small observed cmSASSS, small differences between observed and expected values for low fitted values (due to fixed effects), and the higher variance of the NB distribution than the residuals as values approach zero.

Figure A.5.1. A

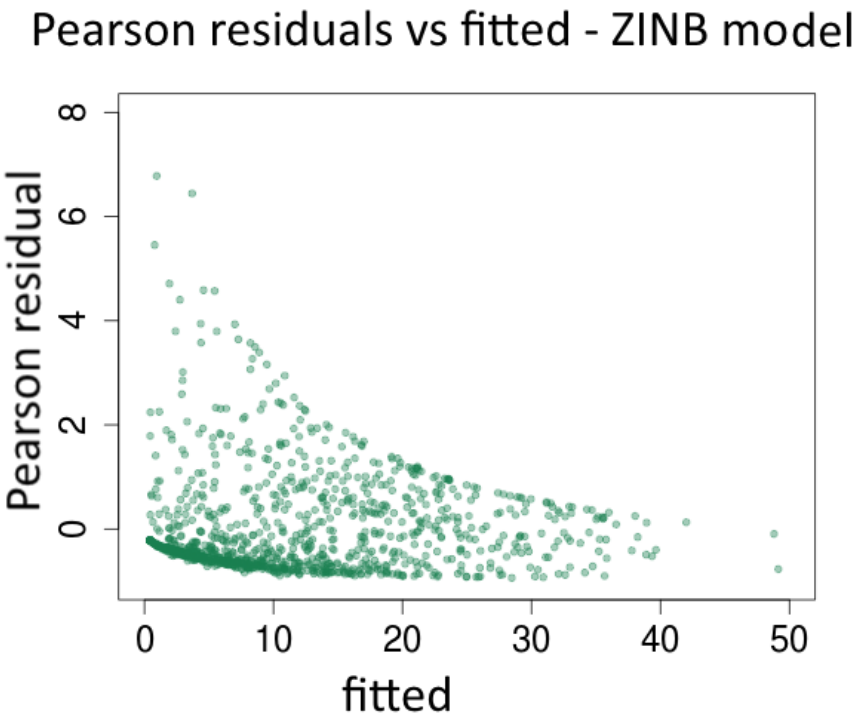


Figure A.5.1. B

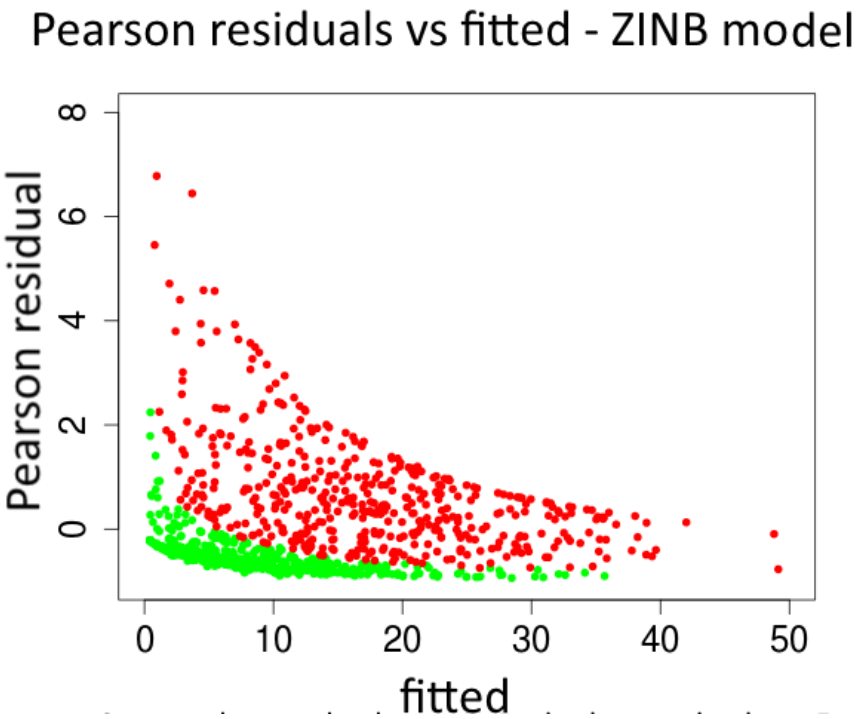


Figure A.5.2 Predicted vs observed plot. This demonstrates that predicted values increase with observed values. Similar to the NB model, predicted values are proportionately lower than observed values.

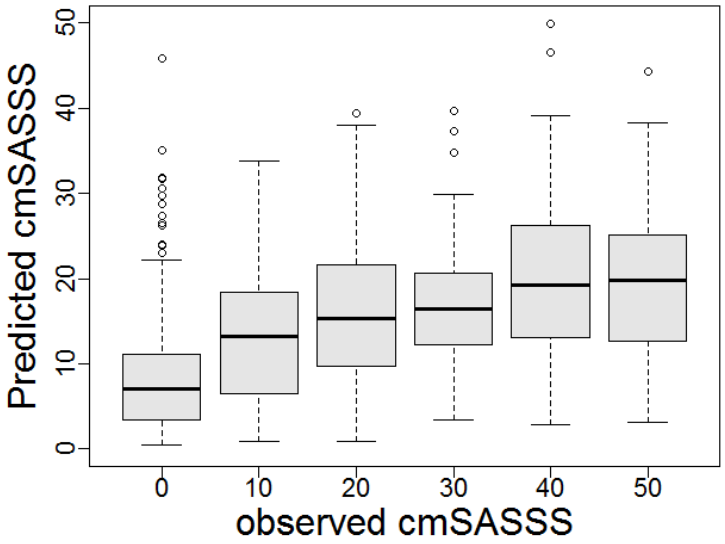


Figure A.5.3 Pearson residual plot. Pearson residuals demonstrate indepence from symptom duration.

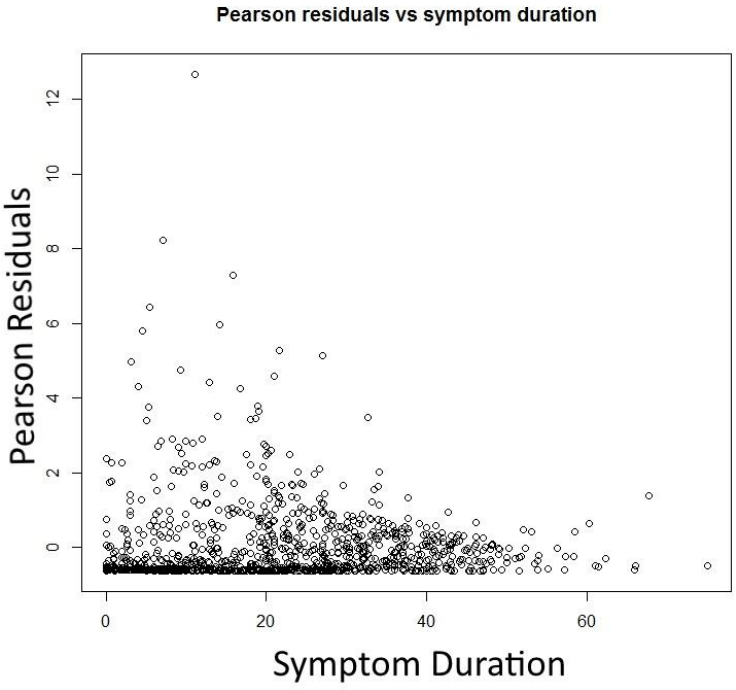
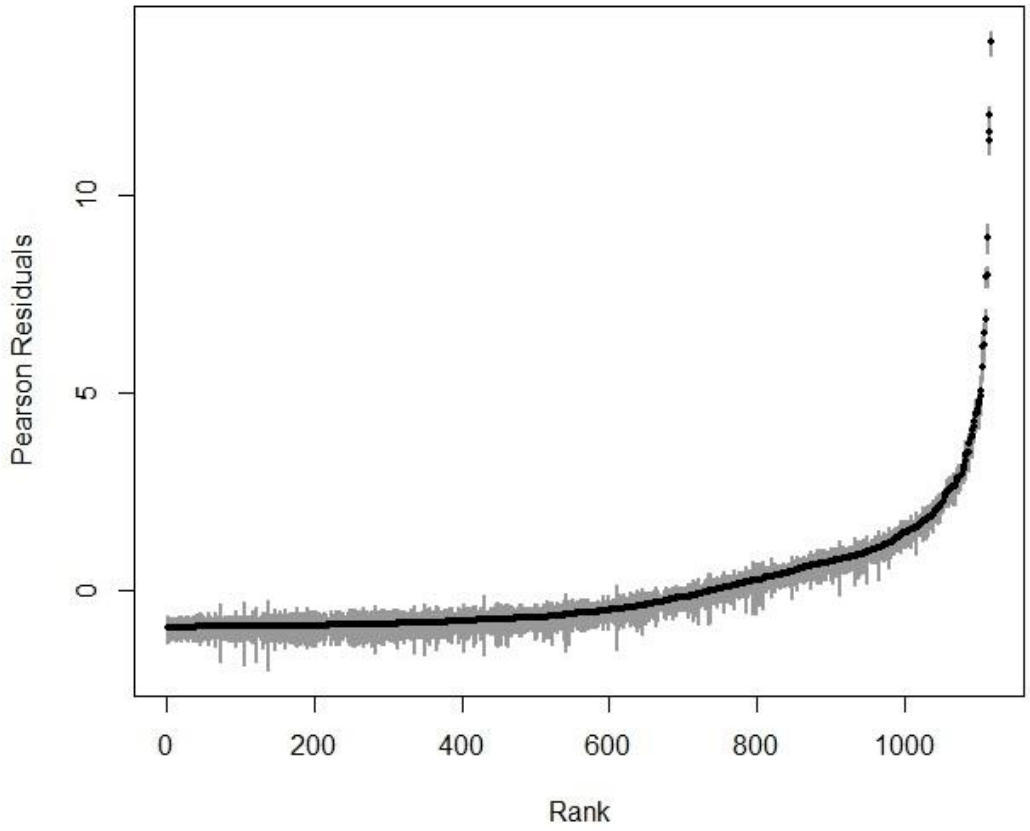


Figure A.5.4 QQ plot of Pearson residuals, with expected residual envelope. Comparison of the residual (black) with the 95% confidence interval estimate (grey) generated by bootstrapping is an established test for ZINB models (114, 115). As the residuals are well contained within the bootstrapped estimate, the model fits adequately and is robust to influential outliers.



A.6 Modelling of SPARCC mSASSS

As per the TASC cohort, the radiographic scores of the SPARCC cohort were approximated by the negative binomial (NB) distribution. The NB model using SPARCC data was overdispersed, similar to the TASC data. Correctible causes of real and apparent overdispersion were investigated. Overdispersion due to excessive counts of zero were addressed by using a ZINB model. The NB and ZINB models were compared using Vuong's test, AIC and goodness-of-fit plots. The ZINB model was superior. Goodness of fit plots for the ZINB model using SPARCC cohort data confirmed an adequate fit, independence of X and Y, and robustness against outliers.

Standard assessment of negative binomial model fit involves examination of standardised 'Pearson residuals' (112, 113).

Overdispersion Calculations

φ = NB model deviance / residual degrees of freedom in NB model = 1.19
$\alpha = 1 / \text{theta}_{\text{NB model}} = 1.29$

NB vs ZINB model comparison

Table A.6.1 Goodness-of-fit tests of NB vs ZINB models. Vuong's test and AIC favour the ZINB model. Lower AIC indicates superior fit.

Model	Vuong's test	AIC
NB	ZINB superior	2055
ZINB	$p < 0.013$	2048

Figure A.6.1 Goodness of fit tests: QQ plot with residual envelope. The QQ plot of Pearson residuals with expected envelope demonstrates adequate fit and robustness to outliers. Comparison of the residual (black) with its 95% confidence interval estimate (generated by bootstrapping) is an established test of model fit for ZINB models (114, 115). As the residuals are well contained within the bootstrapped estimate, the model fits adequately and is robust to influential outliers.

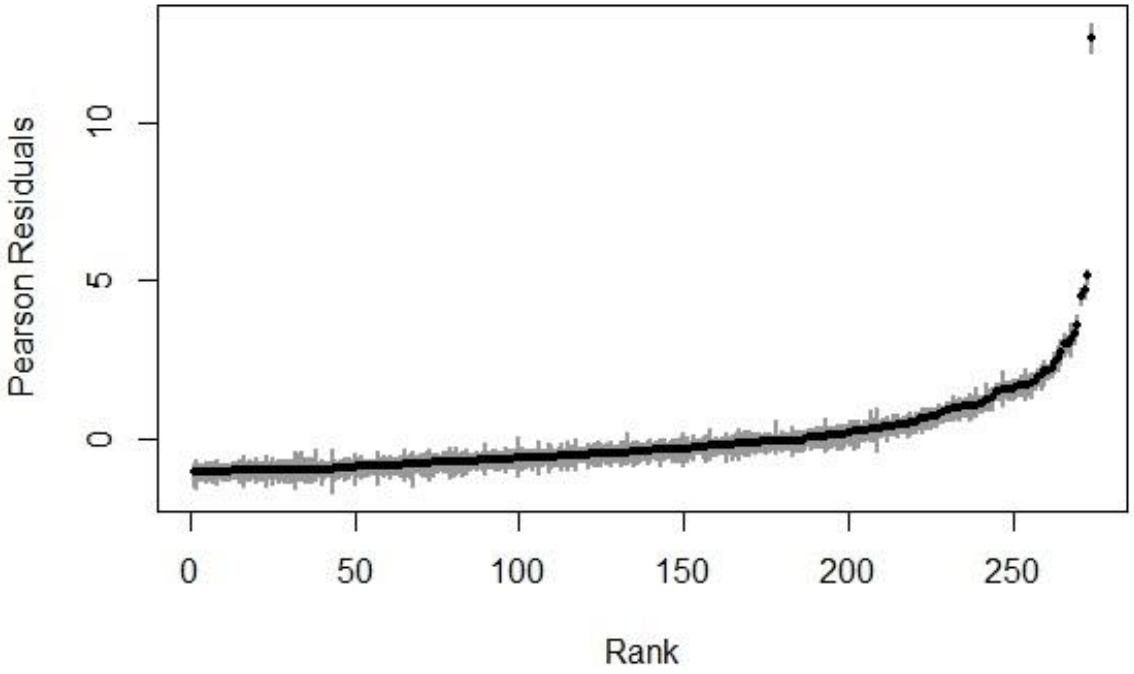


Figure A.6.2 Goodness of fit tests: independence. This plot of Pearson residuals from SPARCC ZINB model. The residuals are independent of symptom duration (the predictor with the largest r^2). This demonstrates that the model adequately adjusts for symptom duration.

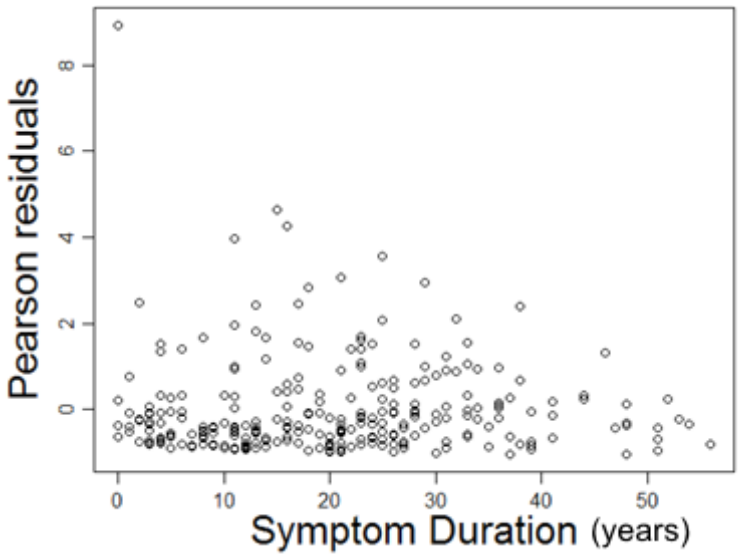


Figure A.6.3 Goodness of fit tests: homogeneity. The Pearson residual vs fitted plot demonstrates adequate homogeneity of the model. There is adequate spread across the residuals for each value of fitted. The distribution of this residual vs fitted plot is as expected for a negative binomial model (See Figure A.4.3, reproduced from Hilbe (113))

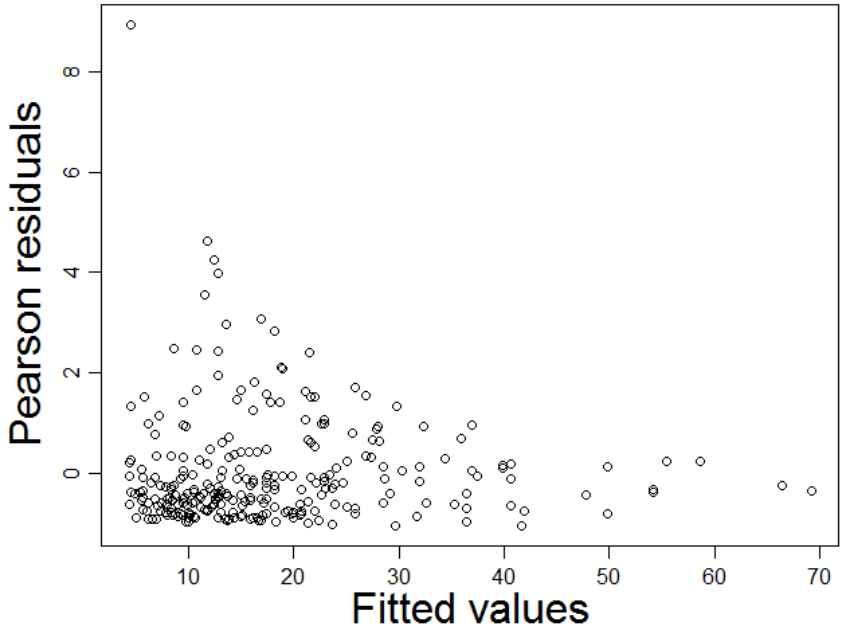


Figure A.6.4 Goodness of fit tests: predicted vs observed plot (SPARCC cohort). Appropriately, predicted mSASSS increases as observed mSASSS increases, however predicted mSASSS is proportionately lower than observed values. Importantly all values are proportionately reduced, preserving the relative differences between subjects. Notably predicted mSASSS plateaus at observed mSASSS over 40. This is an intrinsic deficiency of the negative binomial distribution, where higher scores are less probable, meaning that the model under-predicts the number of subjects with severe disease (See Figure 3.1). The consequence of this deficiency is that subjects with ankylosis will have disproportionately higher residuals, weighting them the most strongly. This is not necessarily an undesirable effect. mSASSS is right censored, where no scores can be higher than 72. The disproportionate weighting of individuals with severe disease may offset the effect of no mSASSS being greater than 72. Another notable feature of the predicted vs observed plot is the fact that for an observed mSASSS of 0, the predicted value is 10. This occurs because the model is describing an average subject - the majority of subjects are male smokers. Addition of fixed effects for the male gender and a smoking history makes the predicted value >0, even though many males have an mSASSS of 0. The alternative would be to have a negative fixed effect for females, which would give female non-smokers a negative predicted mSASSS. In the context of this condition, a negative predicted mSASSS cannot exist and is the less favourable choice.

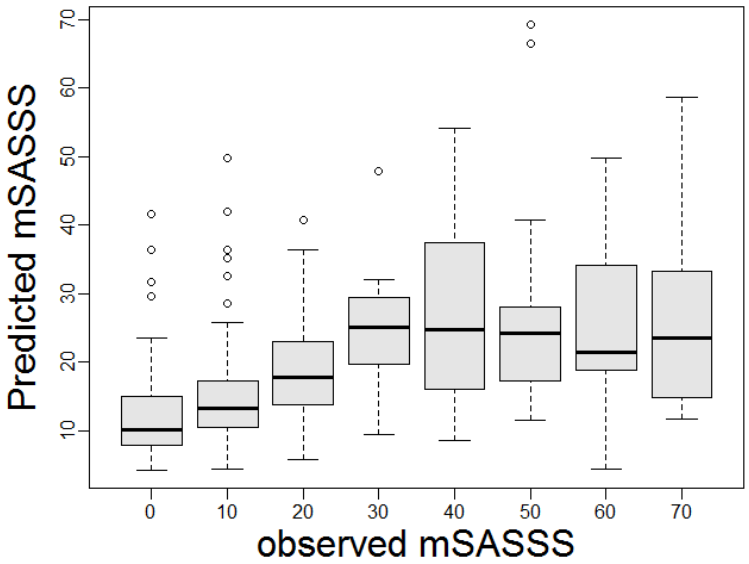


Figure A.6.5 Predicted vs observed plot for TASC cohort. The plateau effect is present but far smaller in the larger TASC model.

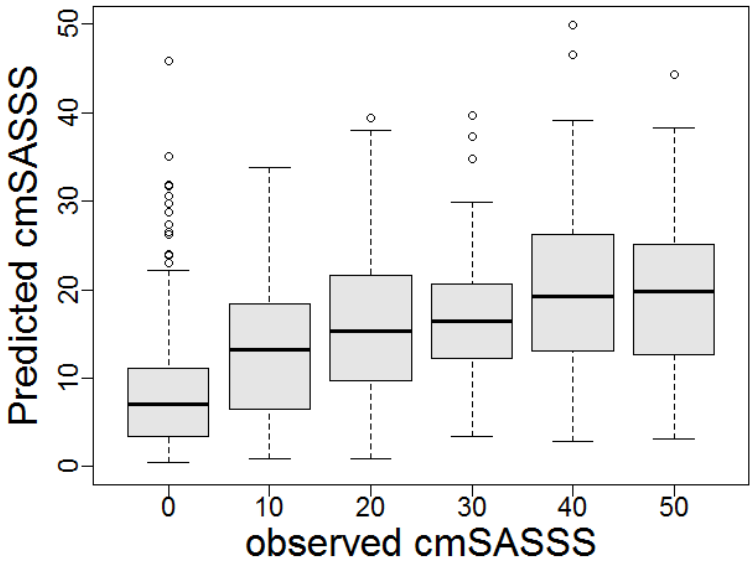
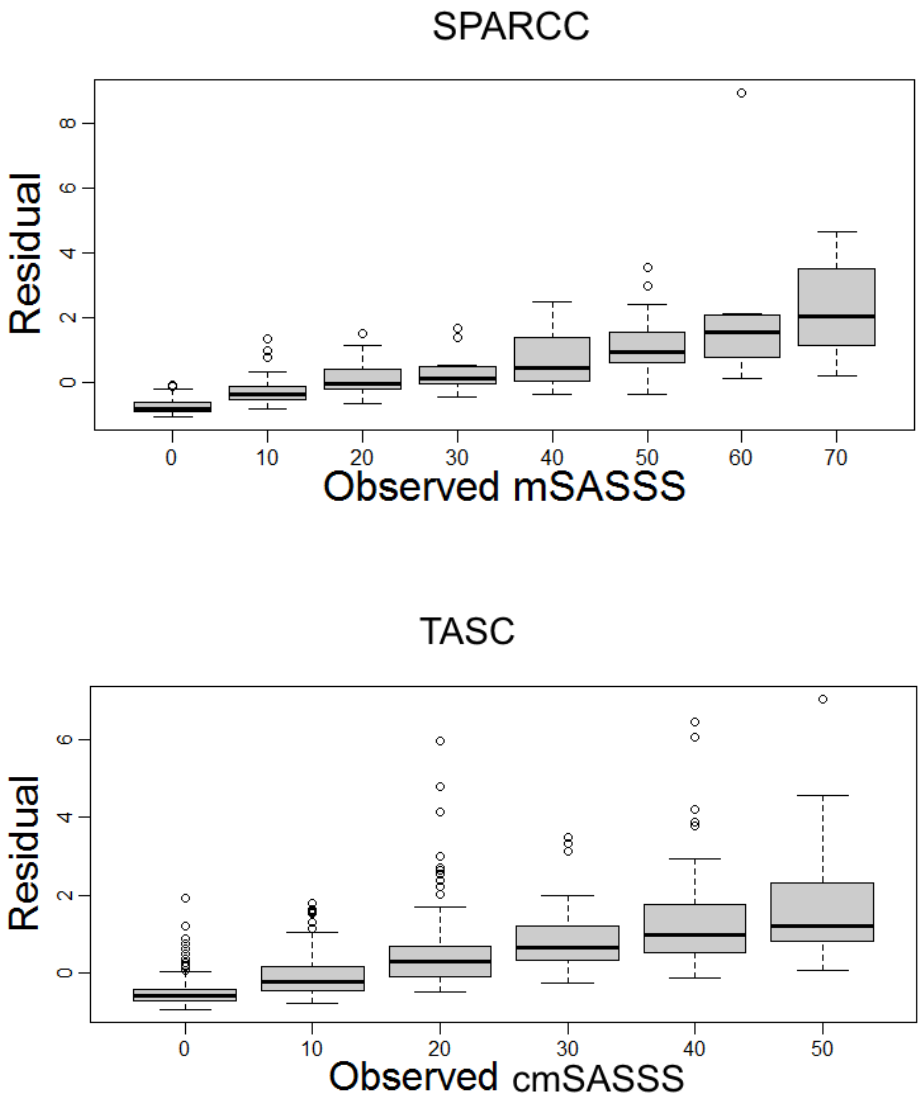


Figure A.6.6 Residual vs Observed Plots for each cohort. In both cohorts, residual values, which were used as a corrected phenotype, increased with observed values. The consequence of the plateau seen in Figure A.6.4 is that the most severe scores in the SPARCC cohort have disproportionately higher residuals, demonstrated by the larger increases between each box once observed mSASSS is greater than 40. The TASC cohort does not display this effect as there is minimal plateauing of predicted values.



References

1. Stolwijk C, Boonen A, van Tubergen A, Reveille JD. Epidemiology of spondyloarthritis. Rheumatic diseases clinics of North America. 2012 Aug;38(3):441-76. PubMed PMID: 23083748.
2. Braun J, Bollow M, Remlinger G, Eggens U, Rudwaleit M, Distler A, et al. Prevalence of spondylarthropathies in HLA-B27 positive and negative blood donors. Arthritis and rheumatism. 1998 Jan;41(1):58-67. PubMed PMID: 9433870. Epub 1998/01/20. eng.
3. Doran MF, Brophy S, MacKay K, Taylor G, Calin A. Predictors of longterm outcome in ankylosing spondylitis. The Journal of rheumatology. 2003 Feb;30(2):316-20. PubMed PMID: 12563688.
4. Brown MA, Kennedy LG, MacGregor AJ, Darke C, Duncan E, Shatford JL, et al. Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and the environment. Arthritis and rheumatism. 1997 Oct;40(10):1823-8. PubMed PMID: 9336417. Epub 1997/10/23. eng.
5. Brown MA, Brophy S, Bradbury L, Hamersma J, Timms A, Laval S, et al. Identification of major loci controlling clinical manifestations of ankylosing spondylitis. Arthritis and rheumatism. 2003 Aug;48(8):2234-9. PubMed PMID: 12905477. Epub 2003/08/09. eng.
6. Brophy S, Hickey S, Menon A, Taylor G, Bradbury L, Hamersma J, et al. Concordance of disease severity among family members with ankylosing spondylitis? The Journal of rheumatology. 2004 Sep;31(9):1775-8. PubMed PMID: 15338499.
7. Stone MA, Pomeroy E, Keat A, Sengupta R, Hickey S, Dieppe P, et al. Assessment of the impact of flares in ankylosing spondylitis disease activity using the Flare Illustration. Rheumatology. 2008 Aug;47(8):1213-8. PubMed PMID: 18539622.
8. Carette S, Graham D, Little H, Rubenstein J, Rosen P. The natural disease course of ankylosing spondylitis. Arthritis and rheumatism. 1983 Feb;26(2):186-90. PubMed PMID: 6600615. Epub 1983/02/01. eng.
9. Sieper J, Appel H, Braun J, Rudwaleit M. Critical appraisal of assessment of structural damage in ankylosing spondylitis: implications for treatment outcomes. Arthritis and rheumatism. 2008 Mar;58(3):649-56. PubMed PMID: 18311819.
10. Aufdermaur M. Pathogenesis of square bodies in ankylosing spondylitis. Annals of the rheumatic diseases. 1989 Aug;48(8):628-31. PubMed PMID: 2782972. Pubmed Central PMCID: 1003836. Epub 1989/08/01. eng.
11. Ball J. Enthesopathy of rheumatoid and ankylosing spondylitis. Annals of the rheumatic diseases. 1971 May;30(3):213-23. PubMed PMID: 4103800. Pubmed Central PMCID: 1005760. Epub 1971/05/01. eng.
12. Maksymowych WP, Chiowchanwisawakit P, Clare T, Pedersen SJ, Ostergaard M, Lambert RG. Inflammatory lesions of the spine on magnetic resonance imaging predict the development of new syndesmophytes in ankylosing spondylitis: evidence of a relationship between inflammation and new bone formation. Arthritis and rheumatism. 2009 Jan;60(1):93-102. PubMed PMID: 19116919.
13. Maksymowych WP, Morency N, Conner-Spady B, Lambert RG. Suppression of inflammation and effects on new bone formation in ankylosing spondylitis: evidence for a window of opportunity in disease modification - CIL. Annals of the rheumatic diseases. 2013 Jan;72(1):23-8. PubMed PMID: 22562977.
14. Baraliakos X, Listing J, Rudwaleit M, Haibel H, Brandt J, Sieper J, et al. Progression of radiographic damage in patients with ankylosing spondylitis: defining the central role of syndesmophytes. Annals of the rheumatic diseases. 2007 Jul;66(7):910-5. PubMed PMID: 17329306. Pubmed Central PMCID: 1955120. Epub 2007/03/03. eng.
15. Lories RJ, Schett G. Pathophysiology of new bone formation and ankylosis in spondyloarthritis. Rheumatic diseases clinics of North America. 2012 Aug;38(3):555-67. PubMed PMID: 23083755.
16. Baraliakos X, Listing J, Haibel H, Sieper J, Braun J. Vertebral Erosions Associated with Spinal Inflammation in Patients with Ankylosing Spondylitis Identified by Magnetic Resonance Imaging: Changes

After 2 Years of Tumor Necrosis Factor Inhibitor Therapy. The Journal of rheumatology. 2013 Aug 1. PubMed PMID: 23908444.
17. Lories RJ, Derese I, de Bari C, Luyten FP. Evidence for uncoupling of inflammation and joint remodeling in a mouse model of spondylarthritis. Arthritis and rheumatism. 2007 Feb;56(2):489-97. PubMed PMID: 17265484.
18. Braem K, Deroose CM, Luyten FP, Lories RJ. Inhibition of inflammation but not ankylosis by glucocorticoids in mice: further evidence for the entheseal stress hypothesis. Arthritis research & therapy. 2012;14(2):R59. PubMed PMID: 22410100. Pubmed Central PMCID: 3446425.
19. Schett G, Stolina M, Dwyer D, Zack D, Uderhardt S, Kronke G, et al. Tumor necrosis factor alpha and RANKL blockade cannot halt bony spur formation in experimental inflammatory arthritis. Arthritis and rheumatism. 2009 Sep;60(9):2644-54. PubMed PMID: 19714640. Epub 2009/08/29. eng.
20. Schett G. Effects of inflammatory and anti-inflammatory cytokines on the bone. European journal of clinical investigation. 2011 Dec;41(12):1361-6. PubMed PMID: 21615394.
21. Goldring SR, Purdue PE, Crotti TN, Shen Z, Flannery MR, Binder NB, et al. Bone remodelling in inflammatory arthritis. Annals of the rheumatic diseases. 2013 Apr;72 Suppl 2:ii52-5. PubMed PMID: 23253928.
22. Cortes A, Hadler J, Pointon JP, Robinson PC, Karaderi T, Leo P, et al. Immunochip - Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. Nature genetics. 2013 Jul;45(7):730-8. PubMed PMID: 23749187. Pubmed Central PMCID: 3757343.
23. Layh-Schmitt G, Yang EY, Kwon G, Colbert RA. HLA-B27 alters the response to tumor necrosis factor alpha and promotes osteoclastogenesis in bone marrow monocytes from HLA-B27-transgenic rats. Arthritis and rheumatism. 2013 Aug;65(8):2123-31. PubMed PMID: 23666508.
24. Hiasa M, Abe M, Nakano A, Oda A, Amou H, Kido S, et al. GM-CSF and IL-4 induce dendritic cell differentiation and disrupt osteoclastogenesis through M-CSF receptor shedding by up-regulation of TNF-alpha converting enzyme (TACE). Blood. 2009 Nov 12;114(20):4517-26. PubMed PMID: 19762488.
25. Sherlock JP, Joyce-Shaikh B, Turner SP, Chao CC, Sathe M, Grein J, et al. IL-23 induces spondyloarthropathy by acting on ROR-gammat+ CD3+CD4-CD8- entheseal resident T cells. Nat Med. 2012 Jul;18(7):1069-76. PubMed PMID: 22772566.
26. Diarra D, Stolina M, Polzer K, Zwerina J, Ominsky MS, Dwyer D, et al. Dickkopf-1 is a master regulator of joint remodeling. Nat Med. 2007 Feb;13(2):156-63. PubMed PMID: 17237793.
27. Heiland GR, Appel H, Poddubnyy D, Zwerina J, Hueber A, Haibel H, et al. High level of functional dickkopf-1 predicts protection from syndesmophyte formation in patients with ankylosing spondylitis. Annals of the rheumatic diseases. 2012 Apr;71(4):572-4. PubMed PMID: 22186710.
28. Kwon SR, Lim MJ, Suh CH, Park SG, Hong YS, Yoon BY, et al. Dickkopf-1 level is lower in patients with ankylosing spondylitis than in healthy people and is not influenced by anti-tumor necrosis factor therapy. Rheumatology international. 2012 Aug;32(8):2523-7. PubMed PMID: 21833531.
29. Daoussis D, Liossis SN, Solomou EE, Tsanaktsi A, Bounia K, Karampetsou M, et al. Evidence that Dkk-1 is dysfunctional in ankylosing spondylitis. Arthritis and rheumatism. 2010 Jan;62(1):150-8. PubMed PMID: 20039407.
30. Uderhardt S, Diarra D, Katzenbeisser J, David JP, Zwerina J, Richards W, et al. Blockade of Dickkopf (DKK)-1 induces fusion of sacroiliac joints. Annals of the rheumatic diseases. 2010 Mar;69(3):592-7. PubMed PMID: 19304568.
31. Chen HA, Chen CH, Lin YJ, Chen PC, Chen WS, Lu CL, et al. Association of bone morphogenetic proteins with spinal fusion in ankylosing spondylitis. The Journal of rheumatology. 2010 Oct;37(10):2126-32. PubMed PMID: 20682677.
32. Maksymowych WP, Landewe R, Conner-Spady B, Dougados M, Mielants H, van der Tempel H, et al. Serum matrix metalloproteinase 3 is an independent predictor of structural damage progression in patients with ankylosing spondylitis. Arthritis and rheumatism. 2007 Jun;56(6):1846-53. PubMed PMID: 17530713.

33. Neidhart M, Baraliakos X, Seemayer C, Zelder C, Gay RE, Michel BA, et al. Expression of cathepsin K and matrix metalloproteinase 1 indicate persistent osteodestructive activity in long-standing ankylosing spondylitis. Annals of the rheumatic diseases. 2009 Aug;68(8):1334-9. PubMed PMID: 18678577.
34. Schett G, Elewaut D, McInnes IB, Dayer JM, Neurath MF. How cytokine networks fuel inflammation: Toward a cytokine-based disease taxonomy. Nat Med. 2013 Jul;19(7):822-4. PubMed PMID: 23836224.
35. Reveille JD, Australo-Anglo-American Spondyloarthritis C, Sims AM, Danoy P, Evans DM, Leo P, et al. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. Nature genetics. 2010 Feb;42(2):123-7. PubMed PMID: 20062062. Pubmed Central PMCID: 3224997. Epub 2010/01/12. eng.
36. Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, Kochan G, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nature genetics. 2011 Aug;43(8):761-7. PubMed PMID: 21743469. Epub 2011/07/12. eng.
37. Poddubnyy D, Rudwaleit M, Haibel H, Listing J, Marker-Hermann E, Zeidler H, et al. Effect of non-steroidal anti-inflammatory drugs on radiographic spinal progression in patients with axial spondyloarthritis: results from the German Spondyloarthritis Inception Cohort. Annals of the rheumatic diseases. 2012 Oct;71(10):1616-22. PubMed PMID: 22459541. Epub 2012/03/31. eng.
38. Kroon F, Landewe R, Dougados M, van der Heijde D. Continuous NSAID use reverts the effects of inflammation on radiographic progression in patients with ankylosing spondylitis. Annals of the rheumatic diseases. 2012 Oct;71(10):1623-9. PubMed PMID: 22532639.
39. Wanders A, Heijde D, Landewe R, Behier JM, Calin A, Olivieri I, et al. Nonsteroidal antiinflammatory drugs reduce radiographic progression in patients with ankylosing spondylitis: a randomized clinical trial. Arthritis and rheumatism. 2005 Jun;52(6):1756-65. PubMed PMID: 15934081. Epub 2005/06/04. eng.
40. Lin Z, Bei JX, Shen M, Li Q, Liao Z, Zhang Y, et al. A genome-wide association study in Han Chinese identifies new susceptibility loci for ankylosing spondylitis. Nature genetics. 2012 Jan;44(1):73-7. PubMed PMID: 22138694.
41. Eskan MA, Jotwani R, Abe T, Chmelar J, Lim JH, Liang S, et al. The leukocyte integrin antagonist Del-1 inhibits IL-17-mediated inflammatory bone loss. Nature immunology. 2012 May;13(5):465-73. PubMed PMID: 22447028. Pubmed Central PMCID: 3330141.
42. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012 Nov 1;491(7422):56-65. PubMed PMID: 23128226. Pubmed Central PMCID: 3498066.
43. Wellcome Trust Case Control C, Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature. 2010 Apr 1;464(7289):713-20. PubMed PMID: 20360734. Pubmed Central PMCID: 2892339. Epub 2010/04/03. eng.
44. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009 Oct 8;461(7265):747-53. PubMed PMID: 19812666. Pubmed Central PMCID: 2831613. Epub 2009/10/09. eng.
45. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nature genetics. 2010 Jul;42(7):565-9. PubMed PMID: 20562875. Pubmed Central PMCID: 3232052. Epub 2010/06/22. eng.
46. Browning SR, Browning BL. Identity-by-descent-based heritability analysis in the Northern Finland Birth Cohort. Human genetics. 2013 Feb;132(2):129-38. PubMed PMID: 23052944. Pubmed Central PMCID: 3543768.
47. Keller MF, Saad M, Bras J, Bettella F, Nicolaou N, Simon-Sanchez J, et al. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. Human molecular genetics. 2012 Nov 15;21(22):4996-5009. PubMed PMID: 22892372. Pubmed Central PMCID: 3576713.
48. Sieper J, Rudwaleit M, Baraliakos X, Brandt J, Braun J, Burgos-Vargas R, et al. The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. Annals of the rheumatic diseases. 2009 Jun;68 Suppl 2:ii1-44. PubMed PMID: 19433414. Epub 2009/05/14. eng.

49. Creemers MC, Franssen MJ, van't Hof MA, Gribnau FW, van de Putte LB, van Riel PL. Assessment of outcome in ankylosing spondylitis: an extended radiographic scoring system. Annals of the rheumatic diseases. 2005 Jan;64(1):127-9. PubMed PMID: 15051621. Pubmed Central PMCID: 1755183. Epub 2004/03/31. eng.
50. Ward MM, Learch TJ, Gensler LS, Davis JC, Jr., Reveille JD, Weisman MH. Regional radiographic damage and functional limitations in patients with ankylosing spondylitis: differences in early and late disease. Arthritis care & research. 2013 Feb;65(2):257-65. PubMed PMID: 23042639. Pubmed Central PMCID: 3541454. Epub 2012/10/09. eng.
51. Wanders AJ, Landewe RB, Spoorenberg A, Dougados M, van der Linden S, Mielants H, et al. What is the most appropriate radiologic scoring method for ankylosing spondylitis? A comparison of the available methods based on the Outcome Measures in Rheumatology Clinical Trials filter. Arthritis and rheumatism. 2004 Aug;50(8):2622-32. PubMed PMID: 15334477.
52. van der Heijde D, Landewe R. Selection of a method for scoring radiographs for ankylosing spondylitis clinical trials, by the Assessment in Ankylosing Spondylitis Working Group and OMERACT. The Journal of rheumatology. 2005 Oct;32(10):2048-9. PubMed PMID: 16206368.
53. Averns HL, Oxtoby J, Taylor HG, Jones PW, Dziedzic K, Dawes PT. Radiological outcome in ankylosing spondylitis: use of the Stoke Ankylosing Spondylitis Spine Score (SASSS). British journal of rheumatology. 1996 Apr;35(4):373-6. PubMed PMID: 8624642. Epub 1996/04/01. eng.
54. Spoorenberg A, de Vlam K, van der Linden S, Dougados M, Mielants H, van de Tempel H, et al. Radiological scoring methods in ankylosing spondylitis. Reliability and change over 1 and 2 years. The Journal of rheumatology. 2004 Jan;31(1):125-32. PubMed PMID: 14705231. Epub 2004/01/06. eng.
55. Baraliakos X, Listing J, Rudwaleit M, Sieper J, Braun J. Development of a radiographic scoring tool for ankylosing spondylitis only based on bone formation: addition of the thoracic spine improves sensitivity to change. Arthritis and rheumatism. 2009 Jun 15;61(6):764-71. PubMed PMID: 19479705. Epub 2009/05/30. eng.
56. Kim TJ, Kim HS, Joo KB, Kim S, Kim TH. Do we really need to evaluate entire cervical spines for squaring score in modified stoke ankylosing spondylitis spinal score? The Journal of rheumatology. 2008 Mar;35(3):477-9. PubMed PMID: 18203314. Epub 2008/01/19. eng.
57. Ramiro A, van Tubergen A, van der Heijde D, Stolwijk C, Dougados M, Van den Bosch F, et al. Erosions and Sclerosis, But Not Squaring, Predict The Development Of New Syndesmophytes: A 12-Year Longitudinal Analysis (OASIS). Arth Rheum. 2013;65(10):S1237.
58. Ramiro S, van Tubergen A, van der Heijde D, Stolwijk C, Bookelman G, Dougados M, et al. Brief report: erosions and sclerosis on radiographs precede the subsequent development of syndesmophytes at the same site: a twelve-year prospective followup of patients with ankylosing spondylitis. Arthritis & rheumatology. 2014 Oct;66(10):2773-9. PubMed PMID: 25048876.
59. van der Heijde D, Landewe R, Baraliakos X, Houben H, van Tubergen A, Williamson P, et al. Radiographic findings following two years of infliximab therapy in patients with ankylosing spondylitis. Arthritis and rheumatism. 2008 Oct;58(10):3063-70. PubMed PMID: 18821688.
60. Wanders A, Landewe R, Spoorenberg A, de Vlam K, Mielants H, Dougados M, et al. Scoring of radiographic progression in randomised clinical trials in ankylosing spondylitis: a preference for paired reading order. Annals of the rheumatic diseases. 2004 Dec;63(12):1601-4. PubMed PMID: 15297280. Pubmed Central PMCID: 1754835.
61. Jang JH, Ward MM, Rucker AN, Reveille JD, Davis JC, Jr., Weisman MH, et al. Ankylosing spondylitis: patterns of radiographic involvement--a re-examination of accepted principles in a cohort of 769 patients. Radiology. 2011 Jan;258(1):192-8. PubMed PMID: 20971774. Pubmed Central PMCID: 3009382. Epub 2010/10/26. eng.
62. Brophy S, Mackay K, Al-Saidi A, Taylor G, Calin A. The natural history of ankylosing spondylitis as defined by radiological progression. The Journal of rheumatology. 2002 Jun;29(6):1236-43. PubMed PMID: 12064842. Epub 2002/06/18. eng.

63. Baraliakos X, Listing J, von der Recke A, Braun J. The natural course of radiographic progression in ankylosing spondylitis--evidence for major individual variations in a large proportion of patients. The Journal of rheumatology. 2009 May;36(5):997-1002. PubMed PMID: 19332632. Epub 2009/04/01. eng.
64. van Tubergen A, Ramiro S, van der Heijde D, Dougados M, Mielants H, Landewe R. Development of new syndesmophytes and bridges in ankylosing spondylitis and their predictors: a longitudinal study. Annals of the rheumatic diseases. 2012 Apr;71(4):518-23. PubMed PMID: 21989544.
65. Ramiro S, Stolwijk C, van Tubergen A, van der Heijde D, Dougados M, van den Bosch F, et al. Evolution of radiographic damage in ankylosing spondylitis: a 12 year prospective follow-up of the OASIS study. Annals of the rheumatic diseases. 2013 Aug 16. PubMed PMID: 23956249.
66. Robinson PC, Brown MA. The genetics of ankylosing spondylitis and axial spondyloarthritis. Rheumatic diseases clinics of North America. 2012 Aug;38(3):539-53. PubMed PMID: 23083754.
67. Boonen A, vander Cruyssen B, de Vlam K, Steinfeld S, Ribbens C, Lenaerts J, et al. Spinal radiographic changes in ankylosing spondylitis: association with clinical characteristics and functional outcome. The Journal of rheumatology. 2009 Jun;36(6):1249-55. PubMed PMID: 19447933.
68. Jaakkola E, Herzberg I, Laiho K, Barnardo MC, Pointon JJ, Kauppi M, et al. Finnish HLA studies confirm the increased risk conferred by HLA-B27 homozygosity in ankylosing spondylitis. Annals of the rheumatic diseases. 2006 Jun;65(6):775-80. PubMed PMID: 16249228. Pubmed Central PMCID: 1798178.
69. Atagunduz P, Aydin SZ, Bahadir C, Erer B, Direskeneli H. Determinants of early radiographic progression in ankylosing spondylitis. The Journal of rheumatology. 2010 Nov;37(11):2356-61. PubMed PMID: 20843901.
70. Kim TJ, Sung IH, Lee S, Joo KB, Choi JH, Park DJ, et al. HLA-B27 homozygosity has no influence on radiographic damage in ankylosing spondylitis: Observation Study of Korean spondyloArthropathy Registry (OSKAR) data. Joint, bone, spine : revue du rhumatisme. 2013 Oct;80(5):488-91. PubMed PMID: 23375452.
71. Ward MM, Hendrey MR, Malley JD, Learch TJ, Davis JC, Jr., Reveille JD, et al. Clinical and immunogenetic prognostic factors for radiographic severity in ankylosing spondylitis. Arthritis Care Res. 2009 Jul 15;61(7):859-66. PubMed PMID: 19565552. Pubmed Central PMCID: 2710412. Epub 2009/07/01. eng.
72. Bartolome N, Szczypiorska M, Sanchez A, Sanz J, Juanola-Roura X, Gratacos J, et al. Genetic polymorphisms inside and outside the MHC improve prediction of AS radiographic severity in addition to clinical variables. Rheumatology. 2012 Aug;51(8):1471-8. PubMed PMID: 22495925. Epub 2012/04/13. eng.
73. Haroon N, Maksymowych WP, Rahman P, Tsui FW, O'Shea FD, Inman RD. Radiographic severity of ankylosing spondylitis is associated with polymorphism of the large multifunctional peptidase 2 gene in the Spondyloarthritis Research Consortium of Canada cohort. Arthritis and rheumatism. 2012 Apr;64(4):1119-26. PubMed PMID: 22034108. Epub 2011/10/29. eng.
74. Wan YI, Shrine NR, Soler Artigas M, Wain LV, Blakey JD, Moffatt MF, et al. Genome-wide association study to identify genetic determinants of severe asthma. Thorax. 2012 Sep;67(9):762-8. PubMed PMID: 22561531.
75. Jensen CJ, Stankovich J, Van der Walt A, Bahlo M, Taylor BV, van der Mei IA, et al. Multiple sclerosis susceptibility-associated SNPs do not influence disease severity measures in a cohort of Australian MS patients. PloS one. 2010;5(4):e10003. PubMed PMID: 20368992. Pubmed Central PMCID: 2848851.
76. International Multiple Sclerosis Genetics C, Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kemppinen A, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nature genetics. 2013 Sep 29. PubMed PMID: 24076602.
77. Patnala R, Clements J, Batra J. Candidate gene association studies: a comprehensive guide to useful in silico tools. BMC genetics. 2013;14:39. PubMed PMID: 23656885. Pubmed Central PMCID: 3655892.
78. Lee W, Reveille JD, Davis JC, Jr., Learch TJ, Ward MM, Weisman MH. Are there gender differences in severity of ankylosing spondylitis? Results from the PSOAS cohort. Annals of the rheumatic diseases. 2007 May;66(5):633-8. PubMed PMID: 17127685. Pubmed Central PMCID: 1954622. Epub 2006/11/28. eng.

79. Haroon N, Inman RD, Learch TJ, Weisman MH, Lee M, Rahbar MH, et al. The Impact of TNF-inhibitors on radiographic progression in Ankylosing Spondylitis. Arthritis and rheumatism. 2013 Jul 1. PubMed PMID: 23818109. Epub 2013/07/03. Eng.
80. Fish EN. The X-files in immunity: sex-based differences predispose immune responses. Nature reviews Immunology. 2008 Sep;8(9):737-44. PubMed PMID: 18728636.
81. Manolagas SC, O'Brien CA, Almeida M. The role of estrogen and androgen receptors in bone health and disease. Nature reviews Endocrinology. 2013 Sep 17. PubMed PMID: 24042328.
82. Poddubnyy D, Haibel H, Listing J, Marker-Hermann E, Zeidler H, Braun J, et al. Cigarette smoking has a dose-dependent impact on progression of structural damage in the spine in patients with axial spondyloarthritis: results from the GErman SPondyloarthritis Inception Cohort (GESPIC). Annals of the rheumatic diseases. 2013 Aug;72(8):1430-2. PubMed PMID: 23625981. Epub 2013/04/30. eng.
83. Chung HY, Machado P, van der Heijde D, D'Agostino MA, Dougados M. Smokers in early axial spondyloarthritis have earlier disease onset, more disease activity, inflammation and damage, and poorer function and health-related quality of life: results from the DESIR cohort. Annals of the rheumatic diseases. 2012 Jun;71(6):809-16. PubMed PMID: 21989541.
84. van der Heijde D, Landewe R, Einstein S, Ory P, Vosse D, Ni L, et al. Radiographic progression of ankylosing spondylitis after up to two years of treatment with etanercept. Arthritis and rheumatism. 2008 May;58(5):1324-31. PubMed PMID: 18438853.
85. Braun J, Baraliakos X, Hermann KG, Deodhar A, van der Heijde D, Inman R, et al. The effect of two golimumab doses on radiographic progression in ankylosing spondylitis: results through 4 years of the GO-RAISE trial. Annals of the rheumatic diseases. 2013 May 3. PubMed PMID: 23644549.
86. van der Heijde D, Salonen D, Weissman BN, Landewe R, Maksymowych WP, Kupper H, et al. Assessment of radiographic progression in the spines of patients with ankylosing spondylitis treated with adalimumab for up to 2 years. Arthritis research & therapy. 2009;11(4):R127. PubMed PMID: 19703304. Pubmed Central PMCID: 2745811.
87. Baraliakos X, Braun J. Response to the eLetter by Maksymowych WP, entitled 'Evidence in Support of the Validity of the TNF Brake Hypothesis'. Annals of the rheumatic diseases. 2013 Sep 20. PubMed PMID: 24058016.
88. Karmakar S, Kay J, Gravallese EM. Bone damage in rheumatoid arthritis: mechanistic insights and approaches to prevention. Rheumatic diseases clinics of North America. 2010 May;36(2):385-404. PubMed PMID: 20510240. Pubmed Central PMCID: 2905601.
89. Visvanathan S, van der Heijde D, Deodhar A, Wagner C, Baker DG, Han J, et al. Effects of infliximab on markers of inflammation and bone turnover and associations with bone mineral density in patients with ankylosing spondylitis. Annals of the rheumatic diseases. 2009 Feb;68(2):175-82. PubMed PMID: 18495735. Pubmed Central PMCID: 2605572.
90. Schoels MM, van der Heijde D, Breedveld FC, Burmester GR, Dougados M, Emery P, et al. Blocking the effects of interleukin-6 in rheumatoid arthritis and other inflammatory rheumatic diseases: systematic literature review and meta-analysis informing a consensus statement. Annals of the rheumatic diseases. 2013 Apr;72(4):583-9. PubMed PMID: 23144446. Pubmed Central PMCID: 3595140.
91. Sieper J, Porter-Brown B, Thompson L, Harari O, Dougados M. Assessment of short-term symptomatic efficacy of tocilizumab in ankylosing spondylitis: results of randomised, placebo-controlled trials. Annals of the rheumatic diseases. 2013 Jun 13. PubMed PMID: 23765873.
92. Poddubnyy D, Haibel H, Listing J, Marker-Hermann E, Zeidler H, Braun J, et al. Baseline radiographic damage, elevated acute-phase reactant levels, and cigarette smoking status predict spinal radiographic progression in early axial spondylarthritis. Arthritis and rheumatism. 2012 May;64(5):1388-98. PubMed PMID: 22127957.
93. Ramiro S, van Tubergen A, van der Heijde D, Stolwijk C, Dougados M, Van den Bosch F, et al. Higher Disease Activity Leads To More Damage In The Early Phases Of Ankylosing Spondylitis: 12-Year Data From The OASIS Cohort. Arthritis and rheumatism. 2013;65(10 (supplement)):S1215.

94. Ramiro S, van der Heijde D, van Tubergen A, Stolwijk C, Dougados M, van den Bosch F, et al. Higher disease activity leads to more structural damage in the spine in ankylosing spondylitis: 12-year longitudinal data from the OASIS cohort. Annals of the rheumatic diseases. 2014 Aug;73(8):1455-61. PubMed PMID: 24812292.
95. Ward MM. Predictors of the progression of functional disability in patients with ankylosing spondylitis. The Journal of rheumatology. 2002 Jul;29(7):1420-5. PubMed PMID: 12136900. Epub 2002/07/26. eng.
96. Ward MM, Reveille JD, Learch TJ, Davis JC, Jr., Weisman MH. Occupational physical activities and long-term functional and radiographic outcomes in patients with ankylosing spondylitis. Arthritis and rheumatism. 2008 Jun 15;59(6):822-32. PubMed PMID: 18512723. Pubmed Central PMCID: 2727685. Epub 2008/06/03. eng.
97. Ramiro S, Van Tubergen A, Landewe R, Boonen A, Stolwijk C, Dougados M, et al. A Physically Demanding Job May Amplify the Effect of Disease Activity on the Development of Syndesmophytes in Patients with AS. Annals of the rheumatic diseases. 2014;73(Suppl 2):2.
98. Gensler L, Haroon N, Reveille JD, Learch R, Brown MA, Weisman M, et al. Socioeconomic status predicts radiographic progression in ankylosing spondylitis. EULAR Abstract. 2013.
99. Brophy S, Pavy S, Lewis P, Taylor G, Bradbury L, Robertson D, et al. Inflammatory eye, skin, and bowel disease in spondyloarthritis: genetic, phenotypic, and environmental factors. The Journal of rheumatology. 2001 Dec;28(12):2667-73. PubMed PMID: 11764216. Epub 2002/01/05. eng.
100. Stolwijk C, van Tubergen A, Castillo-Ortiz JD, Boonen A. Prevalence of extra-articular manifestations in patients with ankylosing spondylitis: a systematic review and meta-analysis. Annals of the rheumatic diseases. 2013 Sep 2. PubMed PMID: 23999006. Epub 2013/09/04. Eng.
101. Machado P, Landewe R, Braun J, Baraliakos X, Hermann KG, Hsu B, et al. Ankylosing spondylitis patients with and without psoriasis do not differ in disease phenotype. Annals of the rheumatic diseases. 2013 Jun;72(6):1104-7. PubMed PMID: 23667171. Epub 2013/05/15. eng.
102. Kim TJ, Lee s, Joo KB, Park DJ, Park YW, Lee S, et al. Presence Of Peripheral Arthritis Delays Spinal Radiographic Progression In Ankylosing Spondylitis: Observation Study Of Korean Spondyloarthropathy Registry (OSKAR) Over 5 Years. Arthritis and rheumatism. 2013;65(10):S646.
103. Apel M, Uebe S, Bowes J, Giardina E, Korendowych E, Juneblad K, et al. Variants in RUNX3 contribute to susceptibility to psoriatic arthritis, exhibiting further common ground with ankylosing spondylitis. Arthritis and rheumatism. 2013 May;65(5):1224-31. PubMed PMID: 23401011. Epub 2013/02/13. eng.
104. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nature reviews Genetics. 2013 Sep;14(9):661-73. PubMed PMID: 23917628.
105. Gensler LS, Ward MM, Reveille JD, Learch TJ, Weisman MH, Davis JC, Jr. Clinical, radiographic and functional differences between juvenile-onset and adult-onset ankylosing spondylitis: results from the PSOAS cohort. Annals of the rheumatic diseases. 2008 Feb;67(2):233-7. PubMed PMID: 17604288.
106. Chen HA, Chen CH, Liao HT, Lin YJ, Chen PC, Chen WS, et al. Clinical, functional, and radiographic differences among juvenile-onset, adult-onset, and late-onset ankylosing spondylitis. The Journal of rheumatology. 2012 May;39(5):1013-8. PubMed PMID: 22422495.
107. Ozgocmen S, Ardicoglu O, Kamanli A, Kaya A, Durmus B, Yildirim K, et al. Pattern of disease onset, diagnostic delay, and clinical features in juvenile onset and adult onset ankylosing spondylitis. The Journal of rheumatology. 2009 Dec;36(12):2830-3. PubMed PMID: 19884272.
108. O'Shea FD, Boyle E, Riarh R, Tse SM, Laxer RM, Inman RD. Comparison of clinical and radiographic severity of juvenile-onset versus adult-onset ankylosing spondylitis. Annals of the rheumatic diseases. 2009 Sep;68(9):1407-12. PubMed PMID: 18782793.
109. Park GS, Wong WK, Khanna D, Gold RH, Paulus HE. Examining radiographic outcomes over time. Rheumatology international. 2014 Feb;34(2):271-9. PubMed PMID: 24166210.
110. Zuur A.F. IEN, Smith G.M. Analysing Ecological Data. New York: Springer; 2007.

111. Lim HK, Li WK, Yu PH. Zero-inflated Poisson regression mixture model. Computational Statistics and Data Analysis. 2014;71:151-8.
112. Zuur A.F. IEN, Walker N.J, Saveliev A.A., Smith G.M. Mixed Effect Models and Extensions in Ecology in R. New York: Springer; 2009.
113. Hilbe J. Negative Binomial Regression: Cambridge University Press; 2011.
114. Garay AM HE, Ortega EM, Lachos VH. On estimation and influence diagnostics for zero-inflated negative binomial regression models. Computational Statistics and Data Analysis. 2010;55:1304-18.
115. Atkinson A. Two graphical displays for outlying and influential observations in regression. Biometrika. 1981;68(1):13-20.
116. Cortes A, Maksymowych WP, Wordsworth BP, Inman RD, Danoy P, Rahman P, et al. Association study of genes related to bone formation and resorption and the extent of radiographic change in ankylosing spondylitis. Annals of the rheumatic diseases. 2014 Apr 10. PubMed PMID: 24651623.
117. Maksymowych WP, Learch T, Lambert RG, Ward M, Haroon N, Inman R, et al. Development and validation of the spondyloarthritis radiography module for calibration of readers using the modified Stoke Ankylosing Spondylitis Spine Score. Arthritis care & research. 2014 Jan;66(1):55-62. PubMed PMID: 23926089.
118. Hochberg. Rheumatology. New York: Elsevier; 2011.
119. Banerjee M. Beyond kappa: a review of interrater agreement measures. Canadian Journal of Statistics. 1999;27(1):3-23.
120. Fleiss J, Levin, L., Paik, MC. Statistical Methods for Rates and Proportions. 3rd ed: John Wiley & Sons; 2003.
121. Zeboulon N, Dougados M, Gossec L. Prevalence and characteristics of uveitis in the spondyloarthropathies: a systematic literature review. Annals of the rheumatic diseases. 2008 Jul;67(7):955-9. PubMed PMID: 17962239.
122. Atkins DC, Baldwin SA, Zheng C, Gallop RJ, Neighbors C. A tutorial on count regression and zero-altered count models for longitudinal substance use data. Psychology of addictive behaviors : journal of the Society of Psychologists in Addictive Behaviors. 2013 Mar;27(1):166-77. PubMed PMID: 22905895. Pubmed Central PMCID: 3513584.
123. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna2013.
124. Jackman S ZA, Kleiber C. pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University. Stanford2012.
125. Reveille JD. Genetics of spondyloarthritis--beyond the MHC. Nature reviews Rheumatology. 2012 May;8(5):296-304. PubMed PMID: 22487796. Epub 2012/04/11. eng.
126. Venables W. N. RBD. MASS: Modern Applied Statistics with S. 2002.
127. Achim Zeileis TH. LMTEST. Vienna2013.
128. Rigby R.A. SDM. Generalized additive models for location, scale and shape. 2005.
129. Schroeder MA. Diagnosing and dealing with multicollinearity. West J Nurs Res. 1990;12(2):175-84; discussion 84-7. PubMed PMID: 2321373.
130. Horton NJ, Fitzmaurice GM. Regression analysis of multiple source and multiple informant data from complex survey samples. Statistics in medicine. 2004 Sep 30;23(18):2911-33. PubMed PMID: 15344194.
131. Magnus JR, Vasnev AL. Local sensitivity and diagnostic tests. Econometrics Journal. 2007;10:166-92.
132. Rosenbaum P. Sensitivity Analysis in Observational Studies. Encyclopaedia of Statistics In Behavioural Science. 4. Chichester: John Wiley & Sons; 2005. p. 1809-14.
133. Yoe C. Principles of Risk Analysis. Boca Raton: CRC Press; 2012.
134. Leffondré K, Abrahamowicz M, Siemiatycki J, Rachet B. Modeling smoking history: a comparison of different approaches. Am J Epidemiol. 2002;156(9):813-23. PubMed PMID: 12396999.

135. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutorials in quantitative methods for psychology. 2012;8(1):23-34. PubMed PMID: 22833776. Pubmed Central PMCID: 3402032.
136. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nature reviews Genetics. 2010;11(7):499-511. PubMed PMID: 20517342.
137. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007 Sep;81(3):559-75. PubMed PMID: 17701901. Pubmed Central PMCID: 1950838.
138. Genetic Analysis of Psoriasis C, the Wellcome Trust Case Control C, Strange A, Capon F, Spencer CC, Knight J, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nature genetics. 2010 Nov;42(11):985-90. PubMed PMID: 20953190. Pubmed Central PMCID: 3749730.
139. Abraham G, Inouye M. Fast principal component analysis of large-scale genome-wide data. PloS one. 2014;9(4):e93766. PubMed PMID: 24718290. Pubmed Central PMCID: 3981753.
140. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics. 2006 Aug;38(8):904-9. PubMed PMID: 16862161.
141. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nature genetics. 2012 Aug;44(8):955-9. PubMed PMID: 22820512. Pubmed Central PMCID: 3696580.
142. Evangelou E, Ioannidis JPA. Meta-analysis methods for genome-wide association studies and beyond. Nature reviews Genetics. 2013;14(6):379-89. PubMed PMID: 23657481.
143. Asimit J, Zeggini E. Rare variant association analysis methods for complex traits. Annu Rev Genet. 2010;44:293-308. PubMed PMID: 21047260.
144. de Bakker PIW, Burtt NP, Graham RR, Guiducci C, Yelensky R, Drake JA, et al. Transferability of tag SNPs in genetic association studies in multiple populations. Nature genetics. 2006;38(11):1298-303. PubMed PMID: 17057720.
145. de Bakker PIW, Ferreira MAR, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Human molecular genetics. 2008;17(R2):R122-8. PubMed PMID: 18852200. Pubmed Central PMCID: PMC2782358.
146. Sham PC, Cherny SS, Purcell S, Hewitt JK. Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. American journal of human genetics. 2000 May;66(5):1616-30. PubMed PMID: 10762547. Pubmed Central PMCID: 1378020.
147. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics. 2003 Jan;19(1):149-50. PubMed PMID: 12499305.
148. Evans DM, Purcell S. Power calculations in genetic studies. Cold Spring Harbor protocols. 2012 Jun;2012(6):664-74. PubMed PMID: 22661434.
149. Spencer CC, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS genetics. 2009 May;5(5):e1000477. PubMed PMID: 19492015. Pubmed Central PMCID: 2688469.
150. Need AC, Goldstein DB. Genome-wide tagging for everyone. Nature genetics. 2006;38(11):1227-8. PubMed PMID: 17072295.
151. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. Nature reviews Genetics. 2014;15(5):335-46. PubMed PMID: 24739678.
152. Gondro Cea. Genome-Wide Association Studies and Genomic Prediction. New York: Springer; 2013.
153. Verma SS, de Andrade M, Tromp G, Kuivaniemi H, Pugh E, Namjou-Khales B, et al. Imputation and quality control steps for combining multiple genome-wide datasets. Front Genet. 2014;5:370. PubMed PMID: 25566314. Pubmed Central PMCID: PMC4263197.

154. Gögele M, Minelli C, Thakkinstian A, Yurkiewich A, Pattaro C, Pramstaller PP, et al. Methods for meta-analyses of genome-wide association studies: critical assessment of empirical evidence. Am J Epidemiol. 2012;175(8):739-49. PubMed PMID: 22427610.
155. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010 Sep 1;26(17):2190-1. PubMed PMID: 20616382. Pubmed Central PMCID: 2922887.
156. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010 Sep 15;26(18):2336-7. PubMed PMID: 20634204. Pubmed Central PMCID: 2935401.
157. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. Nucleic acids research. 2005 Jan 1;33(Database issue):D54-8. PubMed PMID: 15608257. Pubmed Central PMCID: 539985.
158. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015. Nucleic acids research. 2014 Oct 28. PubMed PMID: 25352552.
159. Consortium F, the RP, Clst, Forrest AR, Kawaji H, Rehli M, et al. A promoter-level mammalian expression atlas. Nature. 2014 Mar 27;507(7493):462-70. PubMed PMID: 24670764.
160. Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, et al. The UCSC Genome Browser database: 2014 update. Nucleic acids research. 2014 Jan;42(Database issue):D764-70. PubMed PMID: 24270787. Pubmed Central PMCID: 3964947.
161. de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nature genetics. 2006 Oct;38(10):1166-72. PubMed PMID: 16998491. Pubmed Central PMCID: 2670196.
162. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. American journal of human genetics. 2009 Feb;84(2):210-23. PubMed PMID: 19200528. Pubmed Central PMCID: 2668004.
163. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al. Imputing amino acid polymorphisms in human leukocyte antigens. PloS one. 2013;8(6):e64683. PubMed PMID: 23762245. Pubmed Central PMCID: 3675122.
164. Cerhan JR, Berndt SI, Vijai J, Ghesquieres H, McKay J, Wang SS, et al. Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. Nature genetics. 2014 Nov;46(11):1233-8. PubMed PMID: 25261932. Pubmed Central PMCID: 4213349.
165. Goyette P, Boucher G, Mallon D, Ellinghaus E, Jostins L, Huang H, et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. Nature genetics. 2015 Jan 5. PubMed PMID: 25559196.
166. Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, et al. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nature genetics. 2006;38(11):1251-60. PubMed PMID: 17057719.
167. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci U S A. 2012;109(4):1193-8. PubMed PMID: 22223662. Pubmed Central PMCID: PMC3268279.
168. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, et al. Genome partitioning of genetic variation for complex traits using common SNPs. Nature genetics. 2011 Jun;43(6):519-25. PubMed PMID: 21552263. Pubmed Central PMCID: 4295936.
169. St Pourcain B, Haworth CM, Davis OS, Wang K, Timpson NJ, Evans DM, et al. Heritability and genome-wide analyses of problematic peer relationships during childhood and adolescence. Human genetics. 2014 Dec 17. PubMed PMID: 25515860.
170. Yang J, Lee SH, Goddard ME, Visscher PM. Genome-Wide Complex Trait Analysis (GCTA): Methods, Data Analyses,

and Interpretations. In: Genome-wide association studies and genomic prediction2013. xi, 566 pages p.

171. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS genetics. 2013;9(3):e1003348. PubMed PMID: 23555274. Pubmed Central PMCID: PMC3605113.
172. Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. Human molecular genetics. 2009 Sep 15;18(18):3525-31. PubMed PMID: 19553258.
173. Howe LD, Parmar PG, Paternoster L, Warrington NM, Kemp JP, Briollais L, et al. Genetic influences on trajectories of systolic blood pressure across childhood and adolescence. Circulation Cardiovascular genetics. 2013 Dec;6(6):608-14. PubMed PMID: 24200906.
174. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010 Oct 14;467(7317):832-8. PubMed PMID: 20881960. Pubmed Central PMCID: 2955183.
175. Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. Human molecular genetics. 2007 Jan 1;16(1):24-35. PubMed PMID: 17158188. Pubmed Central PMCID: 2278047.
176. Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, Ntzani EE, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nature genetics. 2012 May;44(5):491-501. PubMed PMID: 22504420. Pubmed Central PMCID: 3338864.
177. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nature genetics. 2010 May;42(5):448-53. PubMed PMID: 20418888. Pubmed Central PMCID: 3080600.
178. Ohnishi T, Yamada K, Watanabe A, Ohba H, Sakaguchi T, Honma Y, et al. Ablation of Mrds1/Ofcc1 induces hyper-gamma-glutamyl transpeptidasemia without abnormal head development and schizophrenia-relevant behaviors in mice. PloS one. 2011;6(12):e29499. PubMed PMID: 22242126. Pubmed Central PMCID: 3248446.
179. Blanco S, Sanz-Garcia M, Santos CR, Lazo PA. Modulation of interleukin-1 transcriptional response by the interaction between VRK2 and the JIP1 scaffold protein. PloS one. 2008;3(2):e1660. PubMed PMID: 18286207. Pubmed Central PMCID: 2243017.
180. Consortium E, Consortium EM, Steffens M, Leu C, Ruppert AK, Zara F, et al. Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32. Human molecular genetics. 2012 Dec 15;21(24):5359-72. PubMed PMID: 22949513.
181. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. American journal of human genetics. 2012 Jan 13;90(1):7-24. PubMed PMID: 22243964. Pubmed Central PMCID: 3257326.
182. Medrano LM, Taxonera C, Marquez A, Barreiro-de Acosta M, Gomez-Garcia M, Gonzalez-Artacho C, et al. Role of TNFRSF1B polymorphisms in the response of Crohn's disease patients to infliximab. Human immunology. 2014 Jan;75(1):71-5. PubMed PMID: 24121042.
183. Waschke KA, Villani AC, Vermeire S, Dufresne L, Chen TC, Bitton A, et al. Tumor necrosis factor receptor gene polymorphisms in Crohn's disease: association with clinical phenotypes. The American journal of gastroenterology. 2005 May;100(5):1126-33. PubMed PMID: 15842589.
184. Baraliakos X, Haibel H, Listing J, Sieper J, Braun J. Continuous long-term anti-TNF therapy does not lead to an increase in the rate of new bone formation over 8 years in patients with ankylosing spondylitis. Annals of the rheumatic diseases. 2013 Mar 27. PubMed PMID: 23505240.
185. Kinoshita T, Wang Y, Hasegawa M, Imamura R, Suda T. PYPAF3, a PYRIN-containing APAF-1-like protein, is a feedback regulator of caspase-1-dependent interleukin-1beta secretion. The Journal of biological chemistry. 2005 Jun 10;280(23):21720-5. PubMed PMID: 15817483.
186. Radian AD, de Almeida L, Dorfleutner A, Stehlik C. NLRP7 and related inflammasome activating pattern recognition receptors and their function in host defense and disease. Microbes and infection / Institut Pasteur. 2013 Jul-Aug;15(8-9):630-9. PubMed PMID: 23618810. Pubmed Central PMCID: 3722249.
187. Khare S, Dorfleutner A, Bryan NB, Yun C, Radian AD, de Almeida L, et al. An NLRP7-containing inflammasome mediates recognition of microbial lipopeptides in human macrophages. Immunity. 2012 Mar 23;36(3):464-76. PubMed PMID: 22361007. Pubmed Central PMCID: 3315380.

188. Zambetti LP, Mortellaro A. NLRPs, microbiota, and gut homeostasis: unravelling the connection. The Journal of pathology. 2014 Aug;233(4):321-30. PubMed PMID: 24740681.
189. Kastbom A, Klingberg E, Verma D, Carlsten H, Forsblad-d'Elia H, Wesamaa J, et al. Genetic variants in CARD8 but not in NLRP3 are associated with ankylosing spondylitis. Scandinavian journal of rheumatology. 2013;42(6):465-8. PubMed PMID: 23547871.
190. Chae JJ, Cho YH, Lee GS, Cheng J, Liu PP, Feigenbaum L, et al. Gain-of-function Pyrin mutations induce NLRP3 protein-independent interleukin-1beta activation and severe autoinflammation in mice. Immunity. 2011 May 27;34(5):755-68. PubMed PMID: 21600797. Pubmed Central PMCID: 3129608.
191. Meng G, Zhang F, Fuss I, Kitani A, Strober W. A mutation in the Nlrp3 gene causing inflammasome hyperactivation potentiates Th17 cell-dominant immune responses. Immunity. 2009 Jun 19;30(6):860-74. PubMed PMID: 19501001. Pubmed Central PMCID: 2764254.
192. Duan J, Chung H, Troy E, Kasper DL. Microbial colonization drives expansion of IL-1 receptor 1-expressing and IL-17-producing gamma/delta T cells. Cell host & microbe. 2010 Feb 18;7(2):140-50. PubMed PMID: 20159619. Pubmed Central PMCID: 4048034.
193. Elinav E, Strowig T, Kau AL, Henao-Mejia J, Thaiss CA, Booth CJ, et al. NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. Cell. 2011 May 27;145(5):745-57. PubMed PMID: 21565393. Pubmed Central PMCID: 3140910.
194. Wlodarska M, Thaiss CA, Nowarski R, Henao-Mejia J, Zhang JP, Brown EM, et al. NLRP6 inflammasome orchestrates the colonic host-microbial interface by regulating goblet cell mucus secretion. Cell. 2014 Feb 27;156(5):1045-59. PubMed PMID: 24581500. Pubmed Central PMCID: 4017640.
195. Costello ME, Ciccia F, Willner D, Warrington N, Robinson PC, Gardiner B, et al. Intestinal dysbiosis in ankylosing spondylitis. Arthritis & rheumatology. 2014 Nov 21. PubMed PMID: 25417597.
196. Sun X, Lewandoski M, Meyers EN, Liu YH, Maxson RE, Jr., Martin GR. Conditional inactivation of Fgf4 reveals complexity of signalling during limb bud development. Nature genetics. 2000 May;25(1):83-6. PubMed PMID: 10802662.
197. Sun X, Mariani FV, Martin GR. Functions of FGF signalling from the apical ectodermal ridge in limb development. Nature. 2002 Aug 1;418(6897):501-8. PubMed PMID: 12152071.
198. Bobick BE, Thornhill TM, Kulyk WM. Fibroblast growth factors 2, 4, and 8 exert both negative and positive effects on limb, frontonasal, and mandibular chondrogenesis via MEK-ERK activation. Journal of cellular physiology. 2007 Apr;211(1):233-43. PubMed PMID: 17167778.
199. Kosaka N, Sakamoto H, Terada M, Ochiya T. Pleiotropic function of FGF-4: its role in development and stem cells. Developmental dynamics : an official publication of the American Association of Anatomists. 2009 Feb;238(2):265-76. PubMed PMID: 18792115.
200. Jazwa A, Kucharzewska P, Leja J, Zagorska A, Sierpniowska A, Stepniewski J, et al. Combined vascular endothelial growth factor-A and fibroblast growth factor 4 gene transfer improves wound healing in diabetic mice. Genetic vaccines and therapy. 2010;8:6. PubMed PMID: 20804557. Pubmed Central PMCID: 2939607.
201. Hughes-Fulford M, Li CF. The role of FGF-2 and BMP-2 in regulation of gene induction, cell proliferation and mineralization. Journal of orthopaedic surgery and research. 2011;6:8. PubMed PMID: 21306641. Pubmed Central PMCID: 3044105.
202. van Gastel N, Stegen S, Stockmans I, Moermans K, Schrooten J, Graf D, et al. Expansion of murine periosteal progenitor cells with fibroblast growth factor 2 reveals an intrinsic endochondral ossification program mediated by bone morphogenetic protein 2. Stem cells. 2014 Sep;32(9):2407-18. PubMed PMID: 24989687.
203. Xiao L, Ueno D, Catros S, Homer-Bouthiette C, Charles L, Kuhn L, et al. Fibroblast growth factor-2 isoform (low molecular weight/18 kDa) overexpression in preosteoblast cells promotes bone regeneration in critical size calvarial defects in male mice. Endocrinology. 2014 Mar;155(3):965-74. PubMed PMID: 24424065. Pubmed Central PMCID: 3929728.
204. Grillo L, Greco D, Pettinato R, Avola E, Potenza N, Castiglia L, et al. Increased FGF3 and FGF4 gene dosage is a risk factor for craniosynostosis. Gene. 2014 Jan 25;534(2):435-9. PubMed PMID: 24120895.

205. Rosen V. BMP2 signaling in bone development and repair. Cytokine & growth factor reviews. 2009 Oct-Dec;20(5-6):475-80. PubMed PMID: 19892583.
206. International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009 Aug 6;460(7256):748-52. PubMed PMID: 19571811. Pubmed Central PMCID: 3912837.
207. St Pourcain B, Skuse DH, Mandy WP, Wang K, Hakonarson H, Timpson NJ, et al. Variability in the common genetic architecture of social-communication spectrum phenotypes during childhood and adolescence. Molecular autism. 2014;5(1):18. PubMed PMID: 24564958. Pubmed Central PMCID: 3940728.
208. Benham H, Rehaume LM, Hasnain SZ, Velasco J, Baillet AC, Ruutu M, et al. Interleukin-23 mediates the intestinal response to microbial beta-1,3-glucan and the development of spondyloarthritis pathology in SKG mice. Arthritis & rheumatology. 2014 Jul;66(7):1755-67. PubMed PMID: 24664521.